

DISCLAIMER:

This document does not meet the
current format guidelines of
the Graduate School at
The University of Texas at Austin.

It has been published for
informational use only.

The Dissertation Committee for Nora Grigore certifies that this is the approved version of the
following dissertation:

Problems of Supererogation

Committee:

Jonathan Dancy, Supervisor

Paul Woodruff, Co-Supervisor

Kathleen Higgins

Rebecca Stangl

Problems of Supererogation

by

Nora Grigore, B.A; M.A; Ph.D.

Dissertation

Presented to the Faculty of the Graduate School

of the University of Texas at Austin

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2017

Problems of Supererogation

by

Nora Grigore, PhD

The University of Texas at Austin, 2017

Supervisor: Jonathan Dancy

Co-supervisor: Paul Woodruff

My main thesis is twofold: I will account first for the *difference* between problems of supererogation, and then for their *relatedness*. First, I will claim that it is very likely that there is no *one* problem of supererogation, but rather a family of problems of supererogation. Secondly, I will claim that the problems of supererogation are related due to the fact that they present conflicts between various aspects of supererogation and one particular view about morality, a view that I will call the “legalistic” view of morality or “morality of law”. Therefore, a first step in my approach of the problem of supererogation will be to show that what has been considered as “the” problem of supererogation consists, rather, in a multitude of problems intricately linked. The second and the third chapters of the thesis will be concerned with this enterprise, of differentiating between various kinds of problems of supererogation. A second step will be to show that most problems of supererogation are connected because they are engendered by an opposition to a particular view of morality. This will be the main point of the fourth chapter. My position will be that whenever a moral theory has adopted (explicitly or not) a legalistic view of morality, that theory will become a hostile environment for the concept of supererogation. I will end by proposing some hypotheses about what could be a more welcoming theoretical environment for supererogation, namely by proposing “morality of virtue” as a likely candidate.

Table of Contents

Chapter One: Morality and Supererogation.....	1
1.0. What is Supererogation?.....	1
1.1. What is the Problem of Supererogation?	4
1.2. The Structure of my Approach to Supererogation.....	10
1.3. Urmson about Saints and Heroes.....	16
1.4. Pybus' Anti-supererogationist Position and Virtue	22
1.5. Mill on Beneficence and Morality.....	32
1.5.1. Justice, Morality, Worthiness and Expediency.....	35
1.5.2. Mill's Puzzle: to be or not to be an Anti-supererogationist.....	37
1.5.3. Heyd about Mill.....	44
1.6. Different Questions, Different Views about Morality.....	50
Chapter Two: Supererogation, Consequentialism and Utilitarianism.....	53
2.0. Utilitarianism, Consequentialism and Supererogation.....	53
2.1. Supererogation under Utilitarian Lenses.....	54
2.2. The Problem of Supererogation in Utilitarian Setting.....	64
2.3. Satisficing as the Reconciliatory Solution between Supererogation and Consequentialism: Michael Slote.....	67
2.4. Dreier's Moral Satisficing.....	76
2.5. The Impersonal Trait of Consequentialist Evaluations.....	83
2.6. Scheffler's Agent-centered Prerogative and Making Room for Supererogation	85
2.7. Zimmerman on Supererogation and Sets of Values.....	91
2.8. Conclusions about Utilitarians, Consequentialists and the Problem of Supererogation	94
2.9. Hurka and Schubert on Extending Permissions.....	96
Chapter Three: Reasons and Supererogation	107
3.0. Introduction.....	108
3.1. Reasons and Supererogation.....	113

3.2. The Problem of Supererogation and its General Form in Dancy's Moral Reasons .	115
3.3. Agent-neutral and Agent-relative Reasons.....	120
3.4. The Two Strategies for Responding to the Problem of Supererogation.....	126
3. 5. Dancy's Solution to the Problem of Supererogation.....	128
3.6. Conclusion: Dancy about Supererogation	135
3.7. Thomas Nagel and the Coexistence of Personal and Impersonal Standpoints.....	137
3.8. Nagel on Supererogation.....	146
3.9. Nagel, Rationality and the Limitations of Human Nature.....	151
3.9.1. Nagel's Answer to the Problem of Supererogation.....	155
3.10. Conclusion: Nagel on Supererogation and Its Problems.....	160
3.11. Raz and Supererogation.....	162
3.12. Raz's Solution to the Problem of Supererogation.....	170
Chapter Four: Supererogation, Obligation and Virtue	172
4.1 The Main Points and Possible Objections.....	173
4.2. Problems of Supererogation and What they May Bring to Light.....	176
4.3. Why Does the Good Have to be Followed by an Ought?.....	185
4.4. Law-like Moral Obligation.....	189
4.5. Looking back: Virtue, Law and Morality.....	191
4.6. Morality, Obligation and Supererogation.....	194
4.7. Thomas Nagel, Susan Wolf and a Common Presupposition.....	198
4.8. Supererogation and the Theory of Virtue	201
4.9. The Case Against the Theory of Virtue.....	210
4.10. Supererogation and Virtue: Conclusion.....	215
Chapter Five: Problems of Supererogation: Final Remarks and Future Study.....	218
References:.....	221

Chapter One: Morality and Supererogation

1.0. What is Supererogation?

Intuitively and commonsensically, we usually recognize morally admirable good deeds. If considerable sacrifice is involved, we are usually quick to admit (at least in case of heavy sacrifices) that such sacrifices are not to be demanded from anyone. This is a rough sketch of the moral phenomenon said to be exemplified especially by saints and heroes (even though more humble examples are not excluded). Our admission that this kind of deeds cannot be morally demanded brings other corollaries (or maybe just different formulations); for example, we would presumably be ready to admit that there are some things that can be morally demanded, some things that can be morally required from each of us, but these cases of sacrifice are not among them. Moreover, when someone does a highly valuable good deed that cannot be demanded, it is said that she or he went 'beyond duty'. One also seems to accept without difficulty that such deeds are morally praiseworthy while their omission is not (or not always) blameworthy.

The commonsensical picture of supererogation seems to bring forth several concepts belonging to the arsenal of moral discourse: moral requirement, duty, morally good, moral value, praiseworthy and blameworthy. The usual general characterization of supererogation also goes along these lines:

Supererogatory acts are those which lie 'above and beyond the call of duty'. Such acts characteristically enjoy a very high degree of value, probably more value than any other act available to the agent. (...) actions which is not wrong of the agent not to do. (Dancy 1998, p.173)

Recent books covering the topic of supererogation start by this rough characterization: "It is often said that works of supererogation involve going beyond the call of duty, doing good in a way that transcends the requirement of moral obligation¹" or "Supererogation is the technical term for the class of actions that go 'beyond the call of duty'. Roughly speaking, supererogatory acts are morally good although not (strictly) required."²

However, strict definitions of supererogation will try to take into account more theoretical notions, such as obligation, permission, right and wrong, consequences and intentions. One might suspect that our ready recognition of this moral phenomenon is more intricate and refined than what we can usually describe, and that our usual vocabulary does not keep up with the actual capacities of discrimination. Therefore, technical definitions have been designed trying to take into account more dimensions, such as the intent or consequences of the action. For example, David Heyd (1982) offers in his book the following definition of supererogatory action:

- 1) Neither obligatory nor forbidden
- 2) The omission is not wrong, it does not deserve sanction or criticism
- 3) Morally good by intended consequences and intrinsic value

¹ Mellema (1991, p.3).

² Heyd (2002/2016).

- 4) Done voluntarily for the sake of someone else's good and is thus meritorious

Gregory Mellema (1991) criticizes Heyd's definition and offers his own along these lines:

- 1) Fulfills no moral duty or obligation
- 2) The performance of the supererogatory act is morally praiseworthy
- 3) The omission of the act is not morally blameworthy

This is only a sketch of what it is usually said about supererogation. But one can already guess the kind of problems it raises: first, problems of boundaries and demarcation brought by any rivalry among definitions; second, problems of membership of particular types of actions (is some particular type of moral action supererogatory or foolish? If an action is brave but not very useful, does it count as supererogatory?). Last but not least, problems of existence may appear, as there are authors denying that supererogation exists in the above description, namely that there are morally good actions one is under no obligation to perform, or that one may omit good actions without any kind of blameworthiness. In this view, the commonsensical point is just a manner of speaking without any serious support from the actual moral concepts employed.

In the next section I will explain in more detail which kind of problems are considered to fall under the heading “the puzzle of supererogation” and which ones are not.

1.1. What is the Problem of Supererogation?

It was called a 'puzzle' by Heyd (1982), a 'paradox' by Horgan and Timmons (2010), or was recognized simply as a 'problem'. The 'problem of supererogation' designates not so much a problem within the domain of supererogatory deeds and their characterization, but rather the strange effect supererogation has on other domains, its status as a trouble-maker when attempts are made to include it in larger theories.

Maybe one should say it from the start: the problem of supererogation is not about identifying which good deeds are supererogatory and which are not. Also, it is not about taking sides in the dispute among supererogationists (i.e. the ones maintaining that there are actions rightly described as supererogatory), anti-supererogationists (i.e. the ones maintaining that deeds of saints and heroes do not go 'beyond duty' and that they are, in some sense, morally required) and quasi-supererogationists (i.e. the ones maintaining that heroic and saintly deeds go 'beyond duty' and are praiseworthy, but failing to act supererogatorily *does* imply some degree of blameworthiness).

At first sight, one might suspect that there is no problem at all, probably because supererogation is familiar enough for our pre-theoretical intuitions, and does not appear to be a very different kind of moral enterprise when theoretically considered. One may be under the impression that supererogation is just about morally good deeds – extremely good deeds – but nevertheless just good deeds; therefore, one might suppose, they should have the theoretical status of every other

good deed in whichever moral theoretical frame we choose to account for them. However, complications quickly escalate as soon as one tries, following the venerable steps of Urmson, to fit the good deeds (the ones usually regarded as supererogatory) within the classical division among deontic categories. According to Urmson's (1958) seminal paper "Saints and Heroes", an excellent moral deed cannot usually be regarded as forbidden, therefore the deed can be seen as either obligatory or merely permitted. If one considers that all morally good deeds are obligatory in some sense, then this is an anti-supererogationist position, according to which there are no supererogatory deeds. However, such a position will usually face difficulties in explaining away the commonsensical moral intuition that one can go beyond what is morally required. On the other hand, if one considers that some extraordinary good deeds are merely permitted, not obligatory, then it would appear that these extraordinary good deeds are a matter of moral indifference, for they can be omitted or not, as the agent pleases. I will discuss Urmson's position in more detail in the next section. For now, an approximate description of the problem could be the following: supererogation, in its commonsensical understanding, seems to be the point where it is claimed that *the 'ought' cannot and should not follow the 'good'*. Supererogatory actions are good but they cannot be required or demanded from anyone, at least not in the commonsensical understanding of what we owe to our fellow human beings. Not doing a certain supererogatory action does not seem to be blamable (at least not in paradigmatic cases), therefore the agent does not seem to be under any kind of obligation to act supererogatorily. But *why should it be the case that we cannot demand what we deem to be a good, and sometimes even an excellent action?*

The puzzle of supererogation starts, I believe, with a simple question: if someone needs help and you are able to provide it, is helping a moral duty (or a moral obligation, or something one ought to do)? The voice of commonsensical moral intuition will most likely reply that indeed, it is a moral obligation to help, when able. But the devil is in details, and as soon as enough details are provided to this hypothetical situation, the above unqualified answer will not match our intuitions anymore. For example, if we can provide help by sacrificing our own life, would that be something we ought to do³? The commonsensical view will hesitate to call such extreme sacrifices 'a moral duty' or to regard them as something to be required. Another, related, way of underlining the dormant conflict between these commonsensical moral intuitions, is to contrast the view that there is a duty to help those in need with the equally commonsensical view that there is such a thing as 'going beyond duty' or doing more than is required. Obviously, if each time one helps, that is only a matter of fulfilling a duty, then there should be no such thing as 'going beyond duty'.

I believe that the root of the puzzle of supererogation may be said to consist in the difficulty of harmonizing these two intuitions: that one can do more than it is required and that any good deed *ought* to be done. However, the puzzle itself appears more often formulated as a question, a question assuming already that *there are* actions that go beyond duty, namely: why should it be the case that we cannot morally require what we deem to be a good, and sometimes even an excellent action? David

³ An interesting feature (to be discussed in the next chapter) of this kind of intuition is that the moral obligation can be maintained when stated in the first person, but the intuition breaks down in the second or third person: one may say "I have a duty to help anyone in need", but the statement "you have a duty to help anyone in need" will not enjoy the same plausibility.

Heyd's formulation, for example, is "how can the moral good not be required as a duty".⁴ Other formulations, e.g. Jonathan Dancy's in *Moral Reasons*, underline the difficulty involved in accounting for omissions of good acts that do not seem to deserve to be called "wrong".

My position is that this puzzle has deceiving simplicity. This not to say that it is difficult to answer (even though this might be true in some sense), but to say that *it hides a multiplicity of related problems*. For example, Heyd presents the puzzle in terms of a problematic relation between morally good actions and duty, while Dancy speaks of the absence of a highly valuable action not being wrong – these are significantly different, albeit related terms. In my view, the image conjured by 'the' puzzle of supererogation is that of an intricate network of conceptual connections where 'the problem' can shift easily from one nexus to another. Consequently, my position will be that the expression "the puzzle of supererogation" it is more likely to refer to a closely related family of problems, rather than to a single well-defined problem.

This point of view is strongly suggested not only by 'the' puzzle's wide variety of formulations, but also by the wide variety of purposes authors have when taking upon them to solve it. For example, Urmson, who started the contemporary discussion about supererogation in "Saints and Heroes", considers that the puzzle is meant to point to the need for a different kind of deontic category, beyond the classical three, namely permission, obligation and interdiction.

⁴ Heyd (2002/2016).

For Horgan and Timmons, in "Untying a Knot from the Inside Out: Reflections on the 'Paradox' of Supererogation" (2010), the point of solving the paradox is to discover a new role that moral reasons can play, namely a "non-requiring, moral merit-conferring role" while denying that "the only way in which a moral reason can favor an action is by tending to require that action."

(...) we propose what we call a nonrequiring, "moral merit-conferring" role that a moral reason can play and that makes sense of the very idea of supererogatory acts: acts that are completely morally optional, but good and morally meritorious to perform. (Horgan and Timmons 2010, p.50)

For Dancy, in *Moral Reasons*, a solution to the puzzle amounts to finding the relation between the evaluative and the deontic domains, as the existence of supererogation requires a gap between what is required (i.e. right or wrong, in the deontic domain) and what is good (i.e. what is considered valuable, in the evaluative domain). Susan Wolf (1982) has a different perspective about the role of supererogation: having a theoretical justification for the distance that supererogation presupposes between what is morally good and what is obligatory, is a salutary trait of a moral theory. That is because the distance between what is morally good and what one ought to do would limit a tendency she considers pernicious, namely that of giving too much weight to moral considerations (by making morally good deeds obligatory) at the expense of considerations regarding one's well-being.

These are quite different enterprises but they all invoke supererogation and its puzzles. The general 'misfit' status of supererogation brings along a difficulty in exactly pinpointing one single problem and its significance: it seems to 'disturb'

several respectable moral notions and theories at once. Therefore, the view that I will take is to treat 'the' puzzle of supererogation as a cluster of problems, rather than as a single problem, and to distinguish between their various conceptual shades. This is a methodological choice which will find its justification in the distinctions to be made in the course of the following chapters. It is, however, just a methodological choice, not a fully-fledged thesis (the thesis would be "the problem of supererogation consists, actually, of many related problems") because, lacking a well-known and agreed upon criterion of identity for problems (i.e. how do we know two problems are actually the same problem or not?), I cannot claim to have a verdict. However, for the same reason, namely the lack of a criterion, the opposing view is bound to face the same difficulty. The ostensible heterogeneity of formulations and purposes is, I believe, an argument that makes my approach (i.e. the "many problems" approach) more reasonable than the usual, opposing approach (i.e. the "one problem" approach).

To conclude, the main task of the present research will be to show that the problem of supererogation actually consists of many *related*, but *different* problems, depending upon the theoretical context in which supererogation is taken into consideration.

I think this is a good place to make a couple of terminological choices. I will prefer the terms "problem" or "puzzle" as they seem to have a larger sphere of applicability than "paradox". Authors do speak about "the paradox of supererogation", even though the problem does not have the classical formal structure of a paradox, namely, if one chooses x , then one quickly arrives by logical means at not- x . I presume a looser sense of "paradox" is being used, namely one in which you

arrive at a contradiction, but a contradiction with a commonsensical and widely held belief, not an internal contradiction of a theory. However, in order to not stir any confusion about the claims involved in using the term "paradox", I will not use this term unless it is present in quotation.

To repeat, my position is that one should rather speak about "problems" of supererogation than "the" problem of supererogation. However, a large majority of authors speak about "the" problem or "the" puzzle; I can only speculate about the reason behind this formulation. Perhaps not much thought was given to the matter or perhaps it is just a way of gesturing towards a cluster of problems without any implication of a commitment towards a certain view on the matter of uniqueness *versus* multiplicity. Given the overwhelming majority of authors that express themselves in this manner, I will have to follow it sometimes, especially when describing their position; however, this will be just a manner of speaking.

1.2. The Structure of my Approach to Supererogation

My main thesis is twofold: I will account first for the *difference* between problems of supererogation, and then for their *relatedness*. First, I will claim that it is very likely that there is no *one* problem of supererogation, but rather a family of problems of supererogation. Secondly, I will claim that the problems of supererogation are related due to the fact that they present conflicts between various aspects of supererogation and one particular view about morality, a view that I will

call the “legalistic” view of morality or “morality of law”. Therefore, a first step in my approach of the problem of supererogation will be to show that what has been considered as “the” problem of supererogation consists, rather, in a multitude of problems intricately linked. The second and the third chapters of the thesis will be concerned with this enterprise, of differentiating between various kinds of problems of supererogation. A second step will be to show that some problems of supererogation are, actually, connected because they are engendered by an opposition to a particular view of morality. This will be the main point of the fourth chapter. My position will be that whenever a moral theory has adopted (explicitly or not) a legalistic view of morality, that theory will become a hostile environment for the concept of supererogation. I will end by proposing some hypotheses about what could be a more welcoming theoretical environment for supererogation, by proposing a “morality of virtue”.

The first chapter will be dedicated to describing the problem and its main concepts: supererogation and moral obligation.

I believe that a good place to start the analysis of the problem of supererogation is the very article that started the conversation about supererogation in ethics, namely Urmson's “Saints and Heroes”. Urmson does not use the word “supererogation” (he discusses, instead, about saintly and heroic actions) and does not explicitly formulate a problem of supererogation. However, his discussion of saintly⁵

⁵ Urmson explicitly states that the word “saintly” does not carry any religious meaning in his usage. “Saintly” here only refers to actions that are extraordinarily good from a moral standpoint and which go against most people's inclinations towards self-preservation or well-being – that is, they presuppose a serious sacrifice on the part of the agent. My usage of the term will be partially in line with his, in that I will not employ the word in the religious meaning either, but I will not take it to have some special connection to supposed general inclinations.

and heroic actions is a good starting point because he first recognizes the challenges posed by these kinds of supererogatory actions to classical moral theories (he comments on utilitarianism, Kantianism and intuitionism) and he is the first to notice that these theories need to make changes to accommodate the commonsensical moral fact of saintly and heroic behavior. But Urmson's most important point, I believe, is about a certain sub-class of heroic and saintly action being inadequately subsumed under either the *obligatory* or the *permitted* moral categories. He argues that some heroic and saintly actions cannot be seen as obligatory because "to characterize an act as a duty is so to demand it",⁶ and they cannot be reasonably demanded (I will discuss his argument in more detail in the next section). For example, one cannot demand a soldier to sacrifice his life in order to save his comrades by throwing himself on a grenade (Urmson's example). This kind of action cannot be seen as merely morally permitted either, because they are not indifferent for morality (they are excellent moral deeds). This is, I think, an important move, one that brings to the fore front of the discussion the problem of *permissions* in morality.

As I see it, there are two main issues when connecting moral *permission* and *obligation* with *supererogation*. First, it is a matter of delimitation between the domain of the obligatory *versus* the domain of merely permitted: supererogatory deeds are those that supposedly go 'beyond' what is obligatory, and therefore they go beyond a certain 'threshold' of obligation (what Urmson would call "the minimum positive demands upon one's fellow men"⁷). This threshold is supposed to work as a delimitation between what is obligatory and what is merely permitted, but authors will

⁶ Urmson (1958 p. 214).

⁷ Urmson (1958, p. 214).

sometimes deny that there is such a threshold. Secondly, it is a matter of discernability within the domain of morally permitted actions: oftentimes the problem will be how to distinguish between kinds of morally permissible deeds, i.e. how to distinguish between the ones permitted because they are morally indifferent, from the ones permitted because they cannot be required (even though they are morally excellent). 'The' problem of supererogation will sometimes be identified with the first issue and sometimes with the second.

In the following section I will discuss Urmson's position. For him, the problem of supererogation seems to be about how one can justify that some heroic and saintly deeds are non-obligatory but also not morally indifferent, and so to draw attention to the fact that there is a trove of different kinds moral actions that are negligently gathered under the label "merely permitted". After Urmson's position, I will discuss a direct reply to it by Elizabeth Pybus (1982), for whom the problem of supererogation shifts towards the first issue, namely an issue of delimitation between moral obligation and permission. Her position is a good example of anti-supererogationism because she denies that there is a domain of merely permitted morally excellent deeds; for her, all morally excellent deeds are, in a certain sense, required. After Pybus, in order to illustrate an opposite position and an actual attempt to systematize the domain of moral obligation versus moral permissions, I will discuss Chapter V from Mill's *Utilitarianism*. This is the discussion included in the first chapter.

The second chapter will be entirely dedicated to act-consequentialist and act-utilitarian theories with their respective problems of supererogation. The problem of supererogation, in its various versions, is a matter of interest for several contemporary

utilitarian and consequentialist positions, mainly because supererogation has little room in a theoretical framework where maximization of the good is required. I will discuss Scheffler's (1982), Slote's (1984), Dreier's (2004), and Zimmerman's (1993) attempts to harmonize the commonsensical idea of supererogation with consequentialist/utilitarian frameworks.

The third chapter is concerned with presenting the problem of supererogation in terms of reasons for action, where supererogation is seen from the standpoint of a larger narrative. Dancy (1993), Nagel (1986) and Raz (1975) will be discussed with an emphasis on their view about the problem of supererogation and where it fits in the larger picture.

The fourth and last chapter will bring all these perspectives together: this is where I will try to show that a large majority of problems of supererogation are born out of an opposition to a certain way of seeing morality. As a consequence, they emerge as different but related problems of supererogation.

This chapter structure corresponds, with approximation, to a division one can easily spot when reading about the status of supererogation in ethics. The first two chapters will present mainly conceptions of supererogation that are discussed in classical terms of duty or obligations and permission. The main focus will be the role of obligations and permissions in morality, mainly because supererogation challenges these roles. The third chapter will present conceptions of supererogation in terms of reasons for action. This division of theoretical frameworks is neither exclusive nor

exhaustive, but it largely corresponds to two possible views that one might have about the problem of supererogation. The usual, general formulation of the problem can be

a) How one could explain that acting in a morally excellent manner is sometimes not obligatory?

Or (the negative):

b) How one could explain that not acting in a morally excellent manner is sometimes permitted?

In both these formulations, one has a choice: to concentrate on the last part of the question, namely on being permitted *versus* being obligatory and the dispute emerging from this opposition; or, to concentrate on the earlier part of the question, namely the one referring to *acting or not acting* according to some obligation or permission. This is a different emphasis, the one preferred by the authors who speak in terms of reasons for action. Of course, this does not mean that one emphasis excludes the other. Both sides will consider the entire question and, therefore, the classical moralists will be concerned with what it means to be allowed to act or not to act according to a rule or in light of an ideal; while authors speaking in terms of reasons for action will be concerned with actions that are permitted or obligatory.

To repeat, this division is not a sharp one: there will be utilitarian authors in the second chapter who speak in terms of reasons, for example Dreier (2004). Also, this division is not exhaustive: there might be problems from the family of the problem of supererogation that are comprised in neither of these two frames.⁸

⁸ For example, approaching supererogation from the perspective of deontic logic might issue different problems.

Nevertheless, I have chosen this structure because I think it is a simple and clear way of dividing between possible ways of seeing the problems raised by supererogation.

1.3. Urmson about Saints and Heroes

To recall, Urmson (1958) does not use the word "supererogation" and he does not explicitly formulate a problem of supererogation. However, a large part of his article "Saints and Heroes" is devoted to establishing a distinction between a "minimum of positive demands" of morality (or "basic rules") and "higher flights of morality", the latter being exemplified by heroic and saintly behavior. His aim in making this distinction is to separate between what he sees as morally obligatory (i.e. duties or "basic rules" or "minimum requirement for living together")⁹ and "acts of moral worth" that are non-obligatory (i.e. some saintly and heroic acts). The specific worry that brings him to make this distinction is that some excellent moral acts, like some heroic acts, might be either ignored or deemed obligatory: "I want to explain the difficulty quite frankly, and to explain why I think that we properly recognize morality that goes beyond duty".¹⁰ To explain why a certain kind of heroic action is not, and should not be, deemed obligatory is, actually, to answer the question "Why is it the case that a certain kind of excellent moral deed is not, and should not be, obligatory?" This question is quite close to the commonsensical version of the problem of supererogation, namely, "Why should it be the case that we cannot

⁹ Urmson (1958, p. 209).

¹⁰ Urmson (1958, p. 210).

demand what we deem to be a good, and sometimes even an excellent action?” Therefore, I think it is very plausible to claim that Urmson does speak of some version of the problem of supererogation, even if he does not label it explicitly in this manner.

In order to understand Urmson's view about the problem of supererogation, I believe the context of his inquiry is relevant. His main purpose is to show that a certain category of moral acts are usually disregarded by the major moral theories, which, in this way, might overlook some important facts of morality. Urmson appears to believe that the main culprit for this overlook is the usual threefold classification of moral actions:

Moral philosophers tend to discriminate, explicitly or implicitly, three types of action from the point of view of moral worth. First, they recognize actions that are a duty, or obligatory, or that we ought to perform, treating these terms as approximately synonymous; second, they recognize actions that are right in so far as they are permissible from a moral standpoint and not ruled out by moral considerations, but that are not morally required of us (...); third, they recognize actions that are wrong, that we ought not to do. (Urmson 1958, p.198)

I think it is interesting that Urmson specifies three kinds of characterizations for the first category, namely “a duty, or obligation, or that we ought to perform”. It is true that most authors treat the three terms as “approximately synonymous”. However, if one chooses to distinguish among the three, then this might affect some conceptions of supererogation. Elizabeth Harman (2016), for example, distinguishes between what is obligatory and what one ought to perform, and therefore she thinks that a supererogatory action is not obligatory, but it might be the case that the agent ought to perform it, all things considered. She calls the omission of a moral act a

“morally permissible moral mistake”: something that it is not obligatory and so it is morally permissible, but a moral mistake nevertheless because one ought to do their best in all circumstances, and not doing this is a moral mistake.

Urmson does not claim that all classical theories use this threefold classification (he admits that some theories have a twofold classification, namely obligatory and forbidden). However, it is a widespread classification that he considers to be inadequate:

To my mind this threefold classification, or any classification that is merely a variation on or elaboration of it, is totally inadequate to the facts of morality; any moral theory that leaves room only for such a classification will in consequence also be inadequate. My main task in this paper will be to show the inadequacy of such a classification by drawing attention to two of the types of action that most conspicuously lie outside such a classification; (Urmson 1958, p.199)

The two types of actions that lie outside the classification are certain kinds of saintly and heroic actions. Urmson believes that commonsense morality easily recognizes and distinguishes between three types of saintly and heroic action:

- a) We usually call someone a saint when fulfilling their duty involves an action going against “inclination, desire or self-interest” that “would lead most people not to do it”; also, we usually call someone a hero when fulfilling their duty involves actions that most people would not do because of “terror, fear, or a drive to self-preservation”.¹¹
- b) We also usually call someone a saint when fulfilling their duty is done “*without effort*” even if the action goes against inclination, desire or self-

¹¹ Urmson (1958, p.200).

interest. The same addition is made for the case of heroic action as described for case a).

- c) Both cases of actions done “resisting to desire and self-interest” and “resisting to fear and self-preservation” (i.e. saintly or heroic), done with or without effort, can be also met when there is *no duty* involved. These are, as Urmson puts it “actions that are far beyond the limits of his duty”.¹²

I have two comments about Urmson's division. First, that the conditions he describes for heroic and saintly action, namely going against self-interest or self-preservation, are another hallmark of supererogation. They amount to saying that extraordinary good moral action requires sacrifices on the part of the agent. However, they will not play a further role in Urmson's argumentation about the deontic status of these actions. They serve only as elements of identification for a certain kind of moral deeds.

Secondly, it is clear from this division that only *some* saintly and heroic acts are considered by Urmson supererogatory, i.e. not falling under the category of obligation or duty. This is an important aspect to be kept in mind when arguing about supererogation because some anti-supererogationist authors will argue that saintly and heroic acts are, actually, someone's duty (Pybus (1982) and New (1974), for example). They are not everyone's duty, but they are a duty for specific categories of agents who have special kinds of commitments. Their argument is directed against the claim that the sacrifice of the agent acting supererogatory is too demanding, too burdensome to be required. The typical reply is that some heavy sacrifices are, actually, duties by the

¹² Urmson, (1958, p.201).

standards of commonsense morality, and therefore, a demanding sacrifice cannot be the sole reason why an action cannot be deemed a duty. Urmson's division can be seen as a rejection of this kind of objection against supererogation. He accepts that some very costly sacrifices can be duties (see categories a) and b) above). However, this does not mean that there are no *other* kinds of costly sacrifices, namely the ones that cannot be required and cannot be seen as duties. Both kinds of sacrifices (the ones that are obligatory and the ones that are not) are present in the commonsense moral panoply, and therefore a proper moral theory should be able to account for both.

Urmson goes on to argue that the third kind of heroic and saintly actions cannot be appropriately described as merely permissible or as obligatory, which makes the threefold division inappropriate:

To summarize, I have suggested that the trichotomy of duties, indifferent actions and wrongdoing is inadequate. There are many kinds of action that involve going beyond duty proper, saintly and heroic actions being conspicuous examples of such kinds of action. (Urmson 1958, p.215)

Urmson leaves open the possibility that there are other kinds of action, beside saintly and heroic, that can be said to go beyond duty.¹³ He argues at length that these acts are not only perceived by commonsense morality as non-obligatory, but also that this status can be justified from a theoretical point of view. His justification is that some moral acts are non-compulsory because it is desirable that, to the highest extent possible, our moral activity be practiced without constraint. Some constraints are

¹³ He mentions that any example of “going a second mile” would qualify as an action going beyond duty proper.

necessary for living together as a society, but not all moral deeds are or should be obligatory:

In the case of moral duties we act to some extent under constraint. We have no choice but to apply pressure on each other to conform in these fundamental matters; here moral principles are like public laws rather than like private ideals. But free choice of the better course of action is always preferable to action under pressure, even when the pressure is but moral. When possible, therefore, it is better that pressure should not be applied and that there should be encouragement and commendation for performance rather than outright demand and censure in the event of nonperformance. (...) But, while there is nothing whatever objectionable in the idea of someone's being pressed to carry out such a basic duty as promise-keeping, there is something horrifying in the thought of pressure being brought on him to perform an act of heroism. (Urmson 1958, pp.213-4)

Urmson does not say why he believes “free choice (...) is always preferable to action under pressure” and I will not venture to guess. I will only say that his division between basic duties and “higher flights of morality” is quite similar to Mill's ideas about moral constraints (they will be discussed in detail in Chapter II). The important point for this research is how he sees the problem of supererogation. Even though he does not ask the question explicitly, I believe Urmson responds to an implicit, presupposed opposition to the idea that some excellent moral deeds might be non-obligatory. Therefore, he probably presupposed a question about why some saintly and heroic deeds are not obligatory. But this just one part of the question, because he wants to say that they are neither obligatory nor merely permitted. Being merely morally permitted would mean, for Urmson, that they are morally indifferent. This has to be wrong because they are obviously laudable from a moral point of view. This is why I think Urmson's version of the problem of supererogation is: “How can one

explain that some saintly and heroic acts are neither morally indifferent nor obligatory?”

In the next section the problem of supererogation is seen from a standpoint that purports to contradict Urmson's main tenet, namely that some saintly and heroic actions are not and should not be obligatory.

1.4. Pybus' Anti-supererogationist Position and Virtue

Elizabeth Pybus' position about supererogation is an example not only of a view explicitly criticizing Urmson, but also an example of how surreptitiously one's view about supererogation might change.

In her reply to Urmson, Pybus (1982) seems not to be concerned with the problem of supererogation. This is because her position is explicitly anti-supererogationist, clearly stating that “such acts do not exist”:

Since the conclusion of his argument has led to the widespread use of the term 'acts of supererogation', and since I do not believe that such acts exist, I propose to argue that the actions with which he is concerned not only can, but should, be contained within the traditional classification. (Pybus 1982, p.190)

Therefore, in her case, it would seem that it makes no sense to ask how one could explain that some excellent moral acts are not obligatory, since it appears there are no such moral acts, and that all excellent moral acts are, in some sense, obligatory. It appears that there is nothing to be explained. Nevertheless, in spite of having a

trenchant opposition to supererogation, Pybus shifts the meaning of one essential term in the picture and the problem of supererogation may, in this way, make a comeback. I will explain in more detail below what are some steps in that direction, and the justification for them. The essential point of Pybus' position is, I think, that in order to be able to say that all morally excellent deeds are a matter of obligation or duty, she changes the meaning of "duty". Urmson (1958) was closely following Mill when he was saying that duty is "something to be exacted, like a debt" or that "to characterize an act as a duty is so to demand it".¹⁴ Pybus (1986) explicitly rejects these ways of seeing duty.¹⁵ Even though the alternative is not clear (how else one should see duty?), what Pybus wants to be able to say is that we all *ought* to follow the moral ideals embodied in heroic or saintly action (this would be our *duty* in her sense), even if this is not something that can, properly speaking, be *demand*ed from us.

Therefore, I believe that Pybus' rather peculiar version of the problem of supererogation can be said to be the following: "How could one explain that heroic or saintly action is something we ought to aspire to, even if the action cannot be demanded?" I believe this is the main idea she is trying to support and explain in her "Saints and Heroes" (1982).

According to Pybus, Urmson is not right to claim that supererogation cannot be contained in the traditional classification of actions into "duties, permissible and wrong" ("although such actions are not duties, they are clearly not wrong, and, more significantly, it is inadequate to describe them as permissible, since anyone who

¹⁴ Urmson (1958, p. 214).

¹⁵ In her reply to McGoldrick's criticism.

performs such actions is worthy of moral praise”).¹⁶ Urmson is not right about supererogation simply because the actions described by Urmson as “supererogatory” actually fall, according to Pybus, under the category of obligation. They are, in a certain sense, a duty. To be sure, Pybus is not saying that everyone has an obligation to jump on a grenade (as in Urmson’s example) or to sacrifice themselves. However, she is saying that we have a duty to strive to actually be as morally good as we can:

By this I do not mean that we should do what the saints and heroes do, but that we must recognize that if we consider such actions susceptible of moral praise, we commit ourselves to saying that what leads to the performance of those actions is part of the equipment of the morally good person which we should all try to be. What I am getting at, therefore, is that in praising what lies behind the actions, i.e. dispositions, or, more specifically, particular virtues. (Pybus 1982, pp.196-7)

Pybus believes that when (and if) we commend heroic and saintly acts we usually commend them from a moral point of view (it could be an aesthetic point of view, but that leads to no commitment according to Pybus). Our commendation of heroic or saintly acts as *moral* acts cannot be just a matter of mere contemplation; if the act is *moral*, then more than just contemplation is called for:

For if, in commending those who perform certain sorts of acts, we do not commit ourselves to the view that those people are the sort of people we ought to be, we are not expressing a moral point of view at all. I cannot at the same time say that something is a moral ideal, and feel that I have no sort of obligation to pursue it. (Pybus 1982, p.195)

Therefore, when commending a saintly or heroic act as a *moral* act we thereby commit ourselves to the ideals or values presupposed by those acts. In other words, it

¹⁶ Urmson (1958, p.193).

commits us to being a certain kind of person and, as a consequence, to act in a certain manner.¹⁷

If I endorse someone's behaviour, and regard him as a saint or hero, by which I mean a morally good man of a special sort, then what I am saying or doing in commending him? I would suggest that if my commendation is genuinely moral, then my genuine act of commendation does commit me to saying that this really is how man ought to be. But if I do have a genuine moral view that this is how people ought to be, then I must think that I, and others, ought to live up to this, and regard those who do not as falling short of the moral standard. (Pybus 1982,p.194)

Therefore, our commendation of a heroic or saintly act implies our recognition (explicitly or not) that a certain ideal is worthwhile pursuing in general, not only for me as an observer, but for everyone. That, for Pybus, is the move that brings "ought" and "should" into the picture, because commending a certain type of action is the same with saying that "we should all try to be" a certain kind of person:

To say, therefore, that someone is a saint or a hero, and thereby to express a moral judgment, is to say that that person has succeeded in being what we all ought to be. He is realizing the worthwhile through his actions. (Pybus 1982, p.196)

Therefore, Pybus is not simply saying that what Urmson calls "supererogatory" is actually obligatory. The picture is more complicated than that. What she is saying is that as long as saintly and heroic deeds are considered morally admirable, then they lay out before us a requirement, something expressible by a "should" or "ought", a moral ideal to be actually pursued if it is to be thought as

¹⁷ I think McGoldrick (1984, p.524) puts it best: "Thus, when we praise the saint or hero we do not commit ourselves to saying that all should do similar acts. We commit ourselves only to the view that all should aim at having, or inculcating within themselves, the dispositions and habitual virtues which give rise to and provide the main spring for actions *of that type*."

genuinely moral. Even if Pybus is not maintaining that we have an obligation to act heroically or saintly, but only that we should strive to be as good as we can, the usual, commonsensical objections come to mind: that even pursuing an ideal cannot be a matter of duty and that no one should (morally or prudentially) be justified in demanding heroic or saintly deeds. In a more elaborate form, these commonsensical objections are brought against Pybus by McGoldrick (1984) in her article “Saints and Heroes: A Plea for the Supererogatory”.

McGoldrick’s first point is that it does not seem to be a matter of duty to strive towards a certain (moral) ideal. Even if one accepts that morality involves developing habitual dispositions that lead to commendable acts, there is something strange in maintaining that morality involves this development¹⁸ *as a matter of duty*.

The second point can be derived from the first. One of the consequences of making adherence to high moral values into a duty seems to be that one should constantly put others before herself or himself. This is, again, an unlikely candidate for a requirement.

McGoldrick’s third point is that the status of saintly and heroic deeds seem to more closely resemble the practice of gift-giving than the practice of exacting a debt (the latter being the one associated by Urmson with duty proper).

However, when Pybus replies¹⁹ to all these objections an interesting picture emerges. It turns out that the objections are misguided, according to Pybus, because she does not have in mind the kind of obligation that would make these objections

¹⁸ McGoldrick accuses Pybus here of *petitio principii*; however, I will skip the details of the argument, as they do not contribute to the plausibility of the counter-argument.

¹⁹ In “A Plea for the Supererogatory: A Reply”.

valid. When she says that we all have an obligation or that we all "should" do something, Pybus is appealing to a different understanding of "obligation". We are not told exactly how we are supposed to understand obligation in this new light, but Pybus is adamant in distinguishing it from Urmson's picture of obligation, namely "a duty which is expressible in rules and can be exacted from all"²⁰. She says that if we understand obligation in Urmson's sense, then saintly and heroic deeds are not obligatory. However they may be said to be obligatory in another sense (which remains mysterious apart from examples):

Now, if moral obligations are seen as quasi-legal ones, so that others may expect me to fulfill them, or to punish non-fulfillment, *then beneficence is not obligatory*.²¹ But that seems to require us to see all moral obligations as obligations of justice, and this is perhaps misleading. (...) 'They need our help, so we must give it' is a coherent moral judgment. We are not suppressing, nor need we add, the premise 'They are entitled to it'. (...) But if really good people are benevolent and loving, and exemplify the ideal to which human beings can aspire, then why should we not regard it as an ideal to which we *should* all aspire? (Pybus 1986, pp.530-1)

Therefore, according to Pybus, heroic action is not obligatory if by "obligation" one understands something that can be required by others and which can imply punishment for non-compliance with the requirement.

Pybus does not insist on this point, but her aim in contradicting Urmson seems to be more comprehensive than merely establishing a status for supererogatory deeds. Namely, she seems to aim at a different picture of morality altogether, a picture where the practice of morality cannot be distinguished from the contemplation of moral deeds; if we want to be able to take ourselves seriously about morality, then we

²⁰ Pybus (1986, p.194).

²¹ My italics.

cannot be mere spectators. And this is where moral obligation intervenes, but it is an obligation of cultivating virtue rather than acting in a certain manner or in certain circumstances:

As far as morality goes, though, we are all (barring psychopaths) players, and not merely spectators. We need, therefore, to adopt the techniques of the good people. Because, to do what is morally worth doing, we need virtues, we ought to develop the virtues to enable us to do what is morally worth doing. (Pybus 1986, p.527)

She indicates further that she takes the consequence of this position to be quite far reaching. One of the consequences, for example, is that her position is not about keeping the threefold classical division of moral categories; rather, her position is that the classical deontic categories (however many they may turn out to be) should be accompanied by a whole new background regarding virtues:

The point is, I think, that the apparent inadequacy of that classification arises from the assumption that actions can or should be morally classified on their own, without any reference to the background of the character and motives of the agent. (...) Actions will still be wrong, permissible or obligatory, but an adequate account of morality will not use this classification of actions on its own, without the required account of virtue. (Pybus 1986, p.529)

From her reply to McGoldrick, it would appear that Pybus does not want to simply leave the classical classification as it is, but to modify it in a way different from Urmson's proposal. And this different way is a different way of thinking about obligation altogether. In the next section I will discuss how this affects the status of supererogatory behavior.

To conclude, Pybus' position seems to be more complex than her explicit categorical denial of supererogation would lead us to suspect. She first states that without obligation there can be no morality; and therefore Urmson, by placing heroic action outside obligation, is actually placing supererogation outside morality:

In distinguishing between a morality of duty and a morality of aspiration, Urmson is (as his uneasiness suggests) unnecessarily lowering the concept of duty, and he is, strictly speaking, putting his aspirations or ideals *outside morality altogether*.²² (Pybus 1982, p.195)

However, then she goes on to say that the obligation she believes to be central to morality is not obligation in the "quasi-legalistic" strong sense that Urmson was using. Therefore, instead of extending obligation into the domain of "morality of aspiration" (as one might be tempted to believe from the bellicose expression of her initial position), she seems to claim that a certain weaker sense of obligation is appropriate for an undivided domain of morality.

This might be regarded as a reasonable strategy if one keeps in mind the dilemma presented by supererogatory deeds: on the one hand it seems obvious that one can do more than required, on the other hand making heroic action not required and therefore merely permissible, seems to allow for indifference towards excellent moral behavior. Pybus' solution to the dilemma appears to take the form of a compromise: supererogation is made obligatory and required, thus alleviating fears from the second horn of the dilemma (because it cannot appear as indifferent to the moral pursuit). The argument given for this obligatory status is that one's moral admiration cannot be merely contemplative; it has to result in a practical commitment

²² My italics.

in order to be moral at all. However, this obligation is not of the kind that allows for punishment for omission or of the kind that would justify expectations from others – in order to alleviate worries coming from the first horn of the dilemma (because, in this way, heroic action is not obligatory in the strong sense of “obligatory”).

One obvious problem with this kind of answer is the mysterious nature of such an obligation. Another, less obvious problem, I think, is the way in which Pybus’ position agrees with a common moral intuition while obscuring an important problem. The appeal of Pybus’ position comes, I think, from the fact that it supports our moral intuition that, indeed, we *should* help other people in need, and we *should* be as good as we can possible be. What this kind of formulation obscures, I think, is the deep divide that might run between a first-person point of view and a second- or third-person point of view. If we say “you should be as good as possible” or “they should be as good as possible” it becomes clear that this is demanding something from someone else, as opposed to the situation where I say “I should be as good as possible”. We are not usually inclined to assert the same kind of demandingness to both points of view because such a requirement seems justified only when it comes from me to myself, not when it comes from me to somebody else, or from somebody else towards me. Using the pronoun “we” obscures the divide between my demands for myself and my demands for others because it gathers everyone in a kind of semantic solidarity such that it is hard to distinguish between points of view: is it “our” demand towards “us” or is it “our” demand towards someone else? Of course “we” should be as courageous as possible and “we should” save lives; the problem is: should we save lives by sacrificing *ourselves* or by sacrificing *someone else*? In the

second case, I think, the counter-intuitiveness becomes obvious, but Pybus does not seem to be aware of the this ambiguity present in the employment of "we".

The way in which Pybus chooses to diffuse the tension present in the dilemma (i.e. to solve the problem of supererogation), namely to eliminate the compulsory (as she puts it, "legalistic") aspect of moral obligation, such that supererogation would be "obligatory" without being actually mandatory, has consequences for the way in which Pybus represents the domain of morality as a whole. If obligation is central to morality, as Pybus claims, and the sense of "obligation" is changed, then it is reasonable to assume that the whole moral landscape changes. In order to exemplify a larger philosophical narrative where moral obligation is, indeed, seen "as something to be exacted like a debt", the next section is dedicated to J.S. Mill and his conception about beneficence and duty.

1.5. Mill on Beneficence and Morality

To recall, for Pybus (1982) the problem of supererogation was "How could one explain that heroic or saintly action is something we ought to aspire to, even if the action cannot be demanded?", while for Urmson, the problem consisted in asking: "How can one explain that some saintly and heroic acts are neither morally indifferent nor obligatory?" The main difference between the two positions was the way in which the authors understood the concept of duty. Urmson took duty to be "something to be exacted like a debt", that is to say, something that would have negative consequences

when omitted. While for Pybus this was a “legalistic” conception of duty which was not appropriate for the moral domain. However, Urmson was following a venerable tradition seeing duty the way he did, and he was echoing Mill when he distinguished between “basic duties” and “higher flights of morality”. In this section I will try to determine what the place of supererogation is according to Mill's views about morality, duty and beneficence.

Mill does not use the word “supererogation” and he does not explicitly formulate a problem of supererogation. However, his Chapter V from *Utilitarianism* discusses duty and what can be said to go beyond duty.²³

Utilitarianism, in virtue of its requirement of maximizing the good, is usually seen as a theory not favoring the thought that supererogation might exist: if it is a duty to do your best in any circumstances, one cannot be said to go beyond duty. This general consideration sheds a peculiar light on the case of John Stuart Mill.

In Chapter V, Mill's main concern is to show that, in spite of our intuition to the contrary, Justice is actually a province of the domain of Expediency, i.e. it rests on utility. Mill admits that Justice seems to most people to have a specific imperative character (strict rules, stern demands, punishable transgressions), and, in this respect, it seems to be very much unlike our impression of what is merely useful. He attributes the imperative character associated with the idea of justice to a certain "strong feeling" that one has in matters of justice. The problem of showing that Justice is actually a province of Utility becomes, therefore, the problem of showing that the

²³ Mill will be discussed in the first chapter because his view of moral obligation has something in common with Urmson's. He is not placed in the second chapter because Chapter II comprises mainly contemporary act-utilitarian position.

"strong feeling" associated with the idea of justice does not have to be something inexplicable and *sui generis*, but can be generated by "general laws of our emotional constitution", the workings of which would satisfy the need to protect a certain interest, to preserve something with a high utility value. In short, the problem is to show that the "common attribute or collection of attributes" gathered under the label "justice" could be said to arouse a strong feeling because of an association with something regarded by everyone as extremely useful:

If in everything which men are accustomed to characterise as just or unjust, some one common attribute or collection of attributes is always present, we may judge whether this particular attribute or combination of attributes would be capable of gathering round it a sentiment of that peculiar character and intensity by virtue of the general laws of our emotional constitution, or whether the sentiment is inexplicable, and requires to be regarded as a special provision of Nature. If we find the former to be the case, we shall, in resolving this question, have resolved also the main problem: if the latter, we shall have to seek for some other mode of investigating it. (Mill 1861, p.177)

This being the target, Mill's strategy is to look into the different components of the idea of justice. He finds that one usually speaks of "just" and "unjust" in the case of legal rights, moral rights, attribution of merit, promises or contracts (everything that may involve "breaking faith with any one"), partial and impartial behavior and problems of equality. As all these seem related but nevertheless a heterogeneous bunch, Mill takes a more historical approach in the idea that he might find something common:

Among so many diverse applications of the term justice, which yet is not regarded as ambiguous, it is a matter of some difficulty to seize the mental link which holds them together, and on which the moral sentiment adhering to the term essentially depends. Perhaps, in this

embarrassment, some help may be derived from the history of the word, as indicated by its etymology. (Mill 1861, p.181)

The “mental link” resulting from historical considerations turns out to be the idea of “legal constraint”²⁴ together with the associated threat of a punishment, even if matters of justice are not always matters of law; because, as Mill explains, even if there is no law actually regulating some things we consider unjust, we still have a feeling that some transgressions are wrong and they should be punished.

Once arrived at the idea of wrongness-that-needs-to-be-punished, Mill's journey starts to become relevant for the idea of supererogation and its place within the domain of morality, even if this was not his main point of interest.

1.5.1. Justice, Morality, Worthiness and Expediency

Mill claims that, even though the idea of punishment for wrongdoing is the desired “mental link” connecting various aspects of the idea of justice, this characteristic limits a domain that is larger than the domain of justice strictly. It is something specific to the domain of morality in general, therefore not providing, as Mill complains, *a difference* between *justice* proper and other “*moral obligations*”:

The above is, I think, a true account, as far as it goes, of the origin and progressive growth of the idea of justice. But we must observe, that it

²⁴ P. 181: "Thus the idea of legal constraint is still the generating idea of the notion of justice, though undergoing several transformations before that notion, as it exists in an advanced state of society, becomes complete."

contains, as yet, nothing to distinguish that obligation from moral obligation in general. For the truth is, that the idea of penal sanction, which is the essence of law, enters not only into the conception of injustice, but into that of any kind of wrong. We do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience. (Mill 1861, p.183)

The picture emerging here is that of the domain of moral obligation delimited by the idea of punishment; within this domain there is a distinction between the domain of justice and the domain of "other moral obligations". Mill does not provide on the spot the difference between the two sub-domains but goes on to distinguish this "realm of punishment" from its complementary, something that he calls "Worthiness" and which, we are to understand, does not belong, strictly speaking, to the domain of morality (because it has no expectation of punishment in case of inaction, punishment being considered inappropriate in this case).

This other domain, of Worthiness, is one where there are other kinds of interactions, not pertaining to duty or "moral obligation"; in these cases there is no place for blame or rightful punishments:

There are other things, on the contrary, which we wish that people should do, which we like or admire them for doing, perhaps dislike or despise them for not doing, but yet admit that they are not bound to do; it is not a case of moral obligation; we do not blame them, that is, we do not think that they are proper objects of punishment. (Mill 1861, p.184)

Accordingly, "oughts" and punishments of any kind belong to the domain of right and wrong which is different from the "desirable or laudable". Mill explains that an action

is assigned to one domain or another depending upon our wish to see the agent compelled or exhorted to act:

(...) that we call any conduct wrong, or employ instead, some other term of dislike or disparagement, according as we think that the person ought, or ought not, to be punished for it; and we say that it would be right to do so and so, or merely that it would be desirable or laudable, according as we would wish to see the person whom it concerns, compelled or only persuaded and exhorted, to act in that manner. (Mill 1861, p.184)

This is the realm of Worthiness, for Mill, where no punishment is applicable, and agents can be only persuaded or exhorted; it is different from the domain of morality in general. And, within the domain of morality, Mill still has to account for the difference between justice and general obligation:

This, therefore, being the characteristic difference which marks off, not justice, but morality in general, from the remaining provinces of Expediency and Worthiness; the character is still to be sought which distinguishes justice from other branches of morality. (Mill 1861, p.184)

To summarize, Mill seems to be distinguishing, at this point, between three domains:

- a) justice – the narrowest;
- b) general obligation – together with justice constitutes the whole domain of right and wrong, that is to say of morality, i.e. of what is punishable -if- wrong
- c) Worthiness – the domain of praise where punishment for omission is inappropriate; therefore, situated outside morality

From the characterization given by Mill to the domain of Worthiness, it would seem the perfect fit with the idea of supererogation: nobody is under any obligation to perform this kind of action (it seems to be implied that Mill has in mind a kind of action that has good consequences), their omission is not punishable, someone can be "persuaded or exhorted" but not coerced. However, if one is persuaded to include supererogation in this third category, Mill's classification would also place supererogation *outside* morality. And, to thicken the plot, Mill speaks of beneficence and generosity as categories belonging to moral obligation (b), not to Worthiness. His justification for this is fully discussed in the next section.

1.5.2. Mill's Puzzle: to be or not to be an Anti-supererogationist

The difference between justice and wider moral obligations resembles, for Mill, the one made by ethical writers of his time between perfect and imperfect duties:

Now it is known that ethical writers divide moral duties into two classes, denoted by the ill-chosen expressions, duties of perfect and of imperfect obligation; the latter being those in which, though the act is obligatory, the particular occasions of performing it are left to our choice, as in the case of charity or beneficence, which we are indeed bound to practise, but not towards any definite person, nor at any prescribed time. (Mill 1861, p.185)

If it is a moral duty to practice beneficence ("we are bound to"), and as a *moral* duty, punishment in some form is appropriate for not doing one's duty, then it would seem that Mill is approaching by this stance the anti-supererogationist position or, at least, a quasi-supererogationist position. This is because one might see punishment as a form

of constraint and, if the omission of a beneficent deed is worthy of punishment, then it would mean that the agent is not free to do or not to do the beneficent action. Alternatively, one might say that beneficence is laudable but its omission is blameworthy so it deserves some kind of punishment: this would be a quasi-supererogationist position. But immediately after presenting his readers with this distinction, Mill takes a step back. The trait chosen by Mill to distinguish between perfect and imperfect "obligations" is not so much the liberty to choose the occasion of practicing beneficence as it is the presence of an *opposing right*:

In the more precise language of philosophic jurists, duties of perfect obligation are those duties in virtue of which a correlative right resides in some person or persons; duties of imperfect obligation are those moral obligations which do not give birth to any right. I think it will be found that this distinction exactly coincides with that which exists between justice and the other obligations of morality. (Mill 1861, p.185)

Mill only goes on to underline that no specific person has a right to demand beneficence from someone: "No one has a moral right to our generosity or beneficence, because we are not morally bound to practise those virtues towards any given individual."

Moreover, he dismisses the claims of "mankind" as whole to our beneficence or charity because such a claim would blur the lines between justice and the rest of the moral domain:

For if a moralist attempts, as some have done, to make out that mankind generally, though not any given individual, have a right to all the good we can do them, he at once, by that thesis, includes generosity and beneficence within the category of justice. He is obliged to say, that our utmost exertions are due to our fellow

creatures, thus assimilating them to a debt (...). Wherever there is a right, the case is one of justice, and not of the virtue of beneficence: and whoever does not place the distinction between justice and morality in general where we have now placed it, will be found to make no distinction between them at all, but to merge all morality in justice. (Mill 1861, p.186)

One problem immediately arising from this argument: if beneficence cannot be demanded either by "given individuals" or "mankind", then in what sense is it still a duty, albeit an indirect one? What kind of binding force is Mill talking about when he says that "we are bound" to practice beneficence if it is not the rightful demand of someone or the general 'everyone'?

Mill does address the problem of what exactly it is to have a right. He claims that:

When we call anything a person's right, we mean that he has a valid claim on society to protect him in the possession of it, either by the force of law, or by that of education and opinion. If he has what we consider a sufficient claim, on whatever account, to have something guaranteed to him by society, we say that he has a right to it. If we desire to prove that anything does not belong to him by right, we think this done as soon as it is admitted that society ought not to take measures for securing it to him, but should leave him to chance, or to his own exertions. (Mill 1861, p.189)

Taking this definition of right into account, what would then be Mill's verdict on beneficence? The emerging picture is a rather complicated one, seemingly hesitant between classical supererogationist claims and classical anti-supererogationist stances, but not necessarily incoherent.

Mill wanted a distinction within morality, namely between justice and "other moral obligations". This distinction is supposed to be similar to the one between

perfect and imperfect duties. And up to this point, by making beneficence a duty and claiming that it is something we are "bound to practice" he seemed to have inclined the balance towards anti-supererogationism. However, by grounding the distinction in the presence/absence of a right and by saying that neither a person nor mankind has a right to our beneficence, things become rather complicated because one is confronted with a 'duty' that no one has the right to ask us to perform. Admittedly, Mill objects to the terminology: he says that "duties of perfect and imperfect obligations" is an "ill-chosen expression" but we are left to guess the reason for his discontent. Even if one chooses not to call beneficence a 'duty', Mill still says that it is something "we are bound to practice". The question is: what would be binding on us if not a duty?

To make things even more puzzling, let us consider the following aspect: the domain of morality, where the difference between justice and beneficence is included, appears to presuppose for Mill the idea of some kind of *punishment*, if not by law, then by public opinion or by "own conscience".

For the truth is, that the idea of penal sanction, which is the essence of law, enters not only into the conception of injustice, but into that of any kind of wrong. We do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow-creatures; if not by opinion, by the reproaches of his own conscience. (Mill 1861, p. 184)

Mill insists that the idea of penal sanction ("the essence of law") is associated not only with injustice, but with wrongdoing in general; wrongdoing ought to be punished by law or by opinion of fellows, or by conscience, but it requires a kind of price to be paid, the perpetrator "owes" something, "like a debt":

(...) It is a part of the notion of Duty in every one of its forms, that a person may rightfully be compelled to fulfill it. Duty is a thing which may be exacted from a person, as one exacts a debt. Unless we think that it might be exacted from him, we do not call it his duty. (Mill 1861, p.184)

Consequently, the following appears to be Mill's idea about the domain of morality:

- A) Morality includes the "hard core" of justice and, in addition to that, other moral obligations.
- B) Justice and the other moral obligations (like beneficence) have *in common* the idea of some kind of a punishment for not complying with their demands, and this is why they both are included in "morality" in general – because this is that distinguishing characteristic of the domain of morality.
- C) Justice and other moral obligations (like beneficence) *are different* (and Mill insists upon the requirement of not conflating the two) because justice presupposes a right (of the person asking for justice) while beneficence does not.

Taking this picture into account, what follows when one *fails to be beneficent* is that this person has committed something wrong and therefore should be punished for committing something immoral. However, how is this punishment to be enforced and by whom since this non-virtuous person has not infringed upon any right? Certainly not by law. One might guess, maybe by society, by the disapproval of peers etc. However, according to Mill's definition, the person wronged by the agent's lack of virtue does not have "a valid claim on society to protect him in the possession of it, either by the force of law, or by that of education and opinion." Therefore, society is

not under an obligation to make sure that the non-virtuous become virtuous or act beneficently either by law *or by opinion*. Therefore the disapproval of society for not being beneficent might occur, but it is not morally bound to punish the perpetrator. Because, as Mill claims, if someone does not have a right to beneficence, then "society ought not to take measures for securing it to him, but should leave him to chance, or to his own exertions". Where does this leave us regarding the whole concept of punishment for a wrong, then?

The situation is puzzling because it would seem that a punishment *is* due for the non-virtuous, however the appropriateness of this punishment does not come from the law or public opinion and the ones affected by the lack of virtue (persons or human societies as a whole) cannot demand it (because they do not have a right), even though Mill seem to admit that a wrong *was* committed.

The only way in which I can make sense of Mill's claim might seem a bit *ad hoc*, but it is the only way in which I can see all the pieces of the puzzle fitting with each other. Mill mentions three main ways of being punished for a wrong committed: "if not by law, by the opinion of his fellow-creatures; if not by opinion, by the reproaches of his own conscience." The first two, namely law and opinion of fellow-creatures are excluded by Mill's considerations about rights because he says that, where there is no right, society does not have to intervene *by law or public opinion* to see that action done or a punishment enacted. Only the third remains as a viable option, namely "reproaches of his own conscience". In other words, the only punishment appropriate for the non-virtuous person may come only from himself. This interpretation will be sustained, I think, by further considerations regarding

Mill's strong inclination towards supporting individual freedom (to be discussed in the next section).

The question to be considered here, relative to this interpretation of Mill, is what (if anything) it can bring to light about supererogation. The difficulty of the task consists in the ambiguous status of a “punishment” that comes from within the agent. Is this something we can rightly call “blame” or punishment?

If one considers the answer to be “Yes”, then virtues like beneficence are not properly speaking supererogatory for Mill, because the failure to act *is* actually followed by blame. Mill would rather fit into the category of quasi-supererogation in this interpretation.

If one considers the answer to be “No”, then Mill's position is closer to supererogationists, but I think the more interesting consequence is that this kind of answer raises questions about the binding moral force that is not the force of duty, in the first place; and then questions about how blame or punishment can be different if it comes from myself to myself or if it comes from someone else for myself. One might wonder if Mill's expression “we are indeed bound to practice [beneficence]” does not hide the same ambiguity of the pronoun “we” already discussed for Elizabeth Pybus: should one understand that “we” are “bound” by our own conscience or by others?

The next section discusses in more detail Heyd's opinion about Mill and supererogation.

1.5.3. Heyd about Mill

Heyd (1982) thinks that Mill is simply undecided between anti-supererogationism and supererogationism, even though he seems to be sympathetic towards the latter:

Mill oscillates between a reductionist, essentially anti-supererogationist attitude, and a recognition of meritorious non-obligatory acts. Like Kant, he is faced by some moral beliefs that cannot be justified in the framework of his rigorous theory. But unlike Kant, Mill is surely inclined to supererogationism. (Heyd 1982, p.83)

I will argue in this section that Heyd's argument for seeing Mill as also inclined towards anti-supererogationism does not have much force if one takes into account Mill's final position about morality and utility.

Mill's sympathy towards supererogationism might come, according to Heyd, from Mill's deep commitment towards individualism and freedom. His brief remarks about perfect and imperfect duties, beneficence and generosity, come from *Utilitarianism*, but the sympathetic attitude might come from *On Liberty*:

The broader justification for this supererogationist deviation from the utilitarian principle is outlined in Mill's *On Liberty*, which takes the opposite starting-point to that of *Utilitarianism*. While the principle of utility takes the point of view of the general, overall good whose promotion is the duty of every individual, the principle of liberty puts the emphasis on the rights of the individual and the limits of the legitimate subordination of his interests to the public or general good (...). (Heyd 1982, p.85)

But the two directions, claims Heyd, cannot be reconciled, unless utility is taken to include the value of individual freedom:

Although Mill tries to reconcile these two principles by suggesting that utility should be understood 'in the largest sense' (so as to include considerations of liberty), I think that the two are ultimately incompatible, and hence that Mill cannot claim to be a pure utilitarian. The distinction between the supererogatory and obligatory as is drawn by Mill himself can be justified in terms of utility only on the basis of a very extended sense of 'utility'. (Heyd 1982, p.85)

However, this is not Heyd's main reason for seeing Mill as sometimes taking the anti-supererogationist position. Heyd thinks, first, that by making beneficence an "imperfect duty" Mill cannot regard it as supererogatory; second, there is something corresponding to supererogation in Mill, namely the domain of Worthiness, but this domain is excluded from morality proper, so supererogation would have no place within morality:

[The] distinction between perfect and imperfect duties, which implies a denial of supererogatory acts, is drawn within the sphere of 'morality in general'. But this sphere is marked off by Mill from 'the remaining provinces of Expediency and Worthiness'. (...) So although there is no room for supererogation beyond justice (i.e. in the realm of imperfect duty), there might be room for it 'beyond morality'. (Heyd 1982, p.84)

The main ground for which I do not think this should be a problem for Mill's sympathy for supererogation is to be found in Mill's answer to his own problem, his original problem: the connection between utility and justice. As I have pointed out, his main concern in Chapter V is not morality or supererogation. Mill wants to show that it plausible to regard Justice as a province of Expediency. His way of showing this is to point out that the strong feeling usually associated with justice (but not with

utility in general) comes from two instincts (i.e. "the impulse of self-defense and the feeling of sympathy") naturally associated to something that everyone regards as supremely useful: security. As this is extremely valuable (Mill would probably like to say "useful") to anyone, the strongest feelings (like the desire to retaliate and punish perpetrators) will be associated with any threat to security and, as a consequence, the harshest imperatives and most sacred rules will gather around protecting security. The strong feelings associated with the idea of justice are, therefore, a consequence of very high utility. As the utility decreases for things that are in the vicinity of 'being safe', so does the feeling of urgency; if utility decreases enough, the thirst for retaliation and the feeling of being endangered (i.e. unsafe) might even disappear for cases where something we regard as useful (but not extremely useful) is threatened. And this is how Mill explains the difference between justice and Worthiness, between the domain where a punishment awaits the trespasser and the domain where no such thing seem to be appropriate: *the behaviors that we only praise and exhort are about things that are not important enough for our survival so that we would grant them a protection by stern rules:*

The moral rules which forbid mankind to hurt one another (in which we must never forget to include wrongful interference with each other's freedom) are more vital to human well-being than any maxims, however important, which only point out the best mode of managing some department of human affairs. (...) By merely giving to each other prudential instruction or exhortation, they may gain, or think they gain, nothing: in inculcating on each other the duty of positive beneficence they have an unmistakable interest, but far less in degree: a person may possibly not need the benefits of others; but he always needs that they should not do him hurt.²⁵ Thus the moralities which protect every

²⁵ This is also an additional psychological argument from Mill: I will uphold firmly rules of justice because I might need others to uphold them for me (when in danger myself); but the same argument loses force when it is made about beneficence instead of security: it does not seem

individual from being harmed by others, either directly or by being hindered in his freedom of pursuing his own good, are at once those which he himself has most at heart, and those which he has the strongest interest in publishing and enforcing by word and deed. (...) Now it is these moralities primarily, which compose the obligations of justice. (Mill 1861, p.189)

I believe Mill's solution to his problem is important because it results in a picture where the differences between justice and the other domains ("imperfect duties" and Worthiness) are differences of degree, namely they are distinguished by degrees of utility. They are all provinces of Expediency, starting with Justice as the narrowest sphere and the highest utility value, which sphere can be logically subsumed under Morality (larger sphere, lower degree of utility), which Morality can be logically subsumed under Worthiness, all of them being sub-domains of Expediency. In this picture, there is no real difference in kind between what belongs to Morality and what belongs to Worthiness. Therefore some actions (like beneficence) that do not fall under justice (which means no one has a right to demand them) might deserve some kind of punishment (or remorse) when omitted; some other actions with good consequences will not deserve even a slight punishment when omitted. As Mill makes clear, this also depends on their consequences, i.e. on utility.²⁶ Therefore, there can be no question, in this interpretation, of excluding supererogation from Morality, because

plausible for everyone to say that they need to be beneficent because they themselves might need beneficence someday.

²⁶ Mill about beneficence: "It is noble to be capable of resigning entirely one's own portion of happiness, or chances of it: but, after all, this self-sacrifice must be for some end; it is not its own end; and if we are told that its end is not happiness, but virtue, which is better than happiness, I ask, would the sacrifice be made if the hero or martyr did not believe that it would earn for others immunity from similar sacrifices? (...) All honour to those who can abnegate for themselves the personal enjoyment of life, when by such renunciation they contribute worthily to increase the amount of happiness in the world; but he who does it, or professes to do it, for any other purpose, is no more deserving of admiration than the ascetic mounted on his pillar. He may be an inspiring proof of what men can do, but assuredly not an example of what they should." (p.147)

there is no firmly defined domain of Morality; there are only degrees of utility. Mill himself speaks about a domain of morality because he recognizes a peculiar feeling attached to a certain zone from the domain of Expediency; but that feeling, he claims, is just the subjective expression of a high degree of utility.

Therefore, supererogation might be very well said to inhabit, in Mill's picture, an approximate domain of utility where no punishment is sought for omitting to act towards obtaining utility. This is already a great victory in a Utilitarian environment. Not only is there no maximization of utility required, we are also told by Mill *that one has no right to require it*. With this unusual (for a Utilitarian) provision in place, the domain of supererogation within this frame seems to be firmly established. Even if one drops the complicated considerations about beneficence and punishment from the previous section and considers that, for Mill, beneficence simply is not supererogatory, there is still room for saying that certainly there can be other types of acts considered supererogatory *proper* in Mill's view. No doubt, this would depend on what do we choose to call "beneficence". For example, it is not controversial to include charity among beneficent actions. One can conjecture that it is possible to regard charity as one of those actions which, in Mill's view, nobody has a right to demand, but which, at the same time, can elicit the agent's remorse when omitted (i.e. there is a 'punishment' of sorts for omission, and therefore the act can be said to fulfill a 'moral obligation'). Beyond this kind of act there should be acts that belong to the province of Worthiness, i.e. of high value and showing virtue, acts that cannot be required but only praised and commended. Heroic acts, e.g. of saving other people's lives by placing one's own in danger, cannot be plausibly seen this way, I believe,

because their omission might be plausibly followed by regret or remorse; therefore, Mill's own criterion would not place them exactly outside moral obligation even if they cannot be required. It is also not so clear that they can be called "beneficent" - someone saving one's life is not exactly a "benefactor", but is, rather, a "savior". However, I think that the most plausible candidate for Mill's province of Worthiness is not a certain kind of good action (its omission can almost always be followed by remorse), but a certain kind of virtuous life. A virtuous life (dedicated to finding a cure for a certain disease, for example) cannot be required, but it is admirable and commendable. Also, the question if the omission of a virtuous life would lead to remorse, does not make much sense, and therefore, one can be sure to leave the province of "moral obligation" behind and to enter the domain where supererogation is possible.

1.6. Different Questions, Different Views about Morality

Mill's view about morality and its provinces provided a map where one could establish a place for supererogation (deeds that are admirable and good the omission of which does not deserve any punishment) even if that was not Mill's intention. What we call today supererogatory acts would fit, according to Mill's characterizations of these domains, somewhere in between Moral Obligation and Worthiness, overlapping partially with both. However, the main point in describing Mill's position was to offer an illustration of a classical position regarding moral *obligation in a strong sense*, or,

as Pybus would put it, in a 'legalistic' sense. Mill explicitly states that the concept of moral obligation is just an extension of the idea of justice, which idea of justice is relying, in turn, on the idea of *punishment* for transgressions:

For the truth is, that the idea of penal sanction, which is the essence of law, enters not only into the conception of injustice, but into that of any kind of wrong. We do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience. (Mill 1861, p.183)

Mill's aim was to show that the common sense usually manifests strong feelings and supports the idea of punishment for injustice, because unjust actions are actions going against something of very high utility in a society: life and safety. When the omission of certain moral actions or transgression of certain moral rules will present a smaller loss for society, they tend not to be punished harshly or not to be punished at all. This is a picture of moral obligation where obligation comes in degrees (of utility), but it is nevertheless a picture where, not fulfilling one's obligation brings about deserved punishment and where obligation can be, indeed, "exacted like a debt". Urmson has a similar picture (he calls his own position "utilitarian", even though it is not painted in degrees of utility): there are "basic rules" of morality, essential to the survival of society and, therefore, deemed obligatory as their transgression is punishable; but there are also "higher flights of morality", admirable deeds that are not and *should* not be required. Urmson affirms that

(...) while there is nothing whatever objectionable in the idea of someone's being pressed to carry out such a basic duty as promise keeping, there is something horrifying in the thought of pressure being brought on him to perform an act of heroism. Though the man might feel himself morally called upon to do the deed, it would be a moral

outrage to apply pressure on him to do such a deed as sacrificing his life for others. (Urmson 1958, p. 214)

Urmson does not address the question why such a pressure would be regarded as horrifying. This would have been an important question to answer because it points at some moral limits for what is can be demanded from a moral agent. Nevertheless, even absent such considerations, I believe that one can notice that, by comparing positions regarding the problem of supererogation, a more *general point of view emerges: not only about moral obligation, but about how one sees morality in general.*

Mill depicts the province of Morality as an extension of Law because, he thinks, some kind of punishment should be attached to the omission of a moral duty. This consideration is revealing, not only about his view on moral duty, but also about the way in which morality is seen as a domain governed by law-like rules. One may contrast this conception with Pybus', who rejects this view of obligation a morality as “legalistic”. For her, matters of moral obligation are matters regarding moral ideals and aspirations. Urmson proposes something that appears to be in between: morality has two domains, one of “legalistic” obligations that can be “extracted like a debt”, and another one reserved to “higher flights of morality”, namely aspirations and ideals.

The complete picture is, of course, more complicated than that. Mill also envisaged a domain of aspiration and ideals (i.e. the province of Worthiness), but this appears to be different from Morality. This is what prompts Heyd's accusation that Mill has banished supererogation outside Morality. I have already discussed his

argument and why I think this accusation is not well-founded. My point here is just that the discussion about the problems surrounding supererogation seems to quickly escalate towards discussing *the status of moral obligation and the image of morality in general*. I believe that discussing moral obligation leads to discussions about morality in general because morality is very often seen to be obligation-centered. But this a theme to be fully discussed in Chapter Four.

The next chapter will explore the various problems that supererogation can raise in a consequentialist or utilitarian theoretical framework.

Chapter Two: Supererogation, Consequentialism and Utilitarianism

2.0. Utilitarianism, Consequentialism and Supererogation

Ethical theories, of both utilitarian and deontological persuasion, tend to have difficulties in accommodating the commonsensical intuition that some morally good deeds cannot be required, i.e. in accommodating supererogation. As Heyd (1982) notices, in their "pure but crude forms", the two kinds of ethical theories have difficulties in accommodating supererogation for different reasons: deontological theories because they tend to assume that the domain of the morally good is exhausted by duty (in various forms); utilitarian theories have difficulties because they tend to require uncompromised maximization of the good, not leaving any space for extraordinary good deeds that should not be required (in the first chapter, it was New (1974) illustrating this position). This does not mean, of course, that more refined versions of the two cannot attempt to accommodate supererogation (the degree of their success is another matter).

To 'accommodate' supererogation means that some plausible explanation has to be given within the theory for the commonsensical intuition that morally excellent deeds cannot be required. That is to say, they will attempt an answer the puzzle of

supererogation, the one generally expressed in vague and general terms by the question 'how can one explain that morally excellent deeds cannot be required?'

The purpose of this chapter is to present several ways in which act-utilitarians (and sometimes act-consequentialists) refine their theories in order to accommodate intuitions about supererogation. My aim is to show that the various solutions presented to the puzzle of supererogation are actually solutions of slightly different problems that can be formulated about supererogation. Some of the differences in posing the problem will come from changes in or restrictions on the meanings of key terms, others will come from variations in the theoretical settings adopted in order to accommodate supererogation.

2.1. Supererogation under Utilitarian Lenses

The commonsensical notion of supererogation has, I claim, deceptive clarity and simplicity. This might be because it constitutes only a broad outline of a possible problem: the details are to be filled in by anyone trying to figure out an explanation for the tenets of supererogation. Hurka and Schubert (2012), for example, paint this broad outline in the following manner:

The concept of supererogation has two sides. On one side, a supererogatory act isn't morally required; on the other side, it's somehow better than its alternative, or "beyond" duty in a sense that connotes superiority. (Hurka and Schubert 2012, p.8)

This is not the only characterization in these terms, of the two sides of supererogation.

Here is another similar characterization:

Supererogatory acts have at least two essential features: i) they must be morally optional in the sense of being neither obligatory nor forbidden and ii) they must be in some sense morally superior to some other act that the agent may permissibly do instead. Differing accounts of supererogation typically vary in terms of what they take to be the relevant sense of moral superiority. (Portmore 2003, p.326)

This two-parts view of supererogation seems straightforward. And yet a multitude of questions are left unanswered: how is the threshold established? What makes the supererogatory deed superior? Is it superior because of the sacrifice of the agent? Could one imagine deeds that go beyond duty with little or no sacrifice? Vessel (2010), for example, identifies several possible reasons (proposed by various authors discussing supererogation) for which supererogatory deeds are said to be better: because the action is praiseworthy or because of the risk involved for the agent or because the benefit for others makes it superior.

Most concur that supererogatory action must be morally optional: neither morally obligatory nor morally forbidden. Furthermore, supererogatory action must be in some way especially valuable, more valuable in this way than some competing morally permissible alternative. The sense in which supererogatory action must be more valuable than a competing morally permissible alternative, however, is a matter of rich controversy. Some believe that supererogatory action must be morally better than a competing permissible alternative. Some believe that the performance of supererogatory action confers more moral praiseworthiness upon its agent than would the performance of a morally permissible competitor. Still others believe that supererogatory action must benefit others to a greater extent than a morally permissible alternative. This author is inclined to endorse this latter analytic requirement on the term "supererogatory." (Vessel 2010, p.302)

There are many other features that can be added to the simple two-parts view of supererogation in order to make possible a full theoretical explanation. For example, one feature that does *not* appear in consequentialist and utilitarian discussions of supererogation (as it is not considered significant for this kind of theory) is the intention of the agent: does the agent have to have a good intention in order for the action to count as truly supererogatory? The features that *do* appear in consequentialist discussions of supererogation sometimes fill in the general picture of supererogation, and sometimes transform it; for while sometimes *consequentialist and utilitarian theories adapt themselves to accommodate supererogation, sometimes the concept of supererogation is adapted to fit an utilitarian frame*. To see this process of double influence unfolding, one needs to look at the outline of a consequentialist/ utilitarian theory and notice *which basic traits of these theories come into conflict with which basic traits of the commonsensical concept of supererogation*. This is the task of the present section.

Several authors²⁷ have noticed that the compelling idea at the core of consequentialism is the one underlining a connection between the moral value of an action and the goodness of its outcome. Philippa Foot²⁸ expresses this connection, between what is morally right and good consequences, as “it can never be right to prefer a worse state of affairs to a better”. Scheffler has two formulations, a stronger one in his 1988 introduction to *Consequentialism and Its Critics*, namely “so far as morality is concerned, what people ought to do is to minimize evil and maximize the good” (Scheffler 1988, p.1) and a weaker one in *The Rejection of Consequentialism*:

²⁷ For example, Philippa Foot (1985), Samuel Scheffler (1982, 1988) and Douglas Portmore (2003).

²⁸ Utilitarianism and the Virtues”, p. 227 in Scheffler (ed.).

I believe that utilitarianism refuses to fade from the scene in large part because, as the most familiar consequentialist theory, it is the major recognized normative theory incorporating the deeply plausible-sounding feature that one may always do what would lead to the best available outcome overall. (Scheffler 1982, p.4)

Also, Portmore in “Position-Relative Consequentialism, Agent-Centered Options, and Supererogation” maintains that consequentialist theories adopt “the compelling idea that it can never be wrong to bring about the best state of affairs.”²⁹

Now, the idea that “one may always do what would lead to the best available outcome overall” *per se* does not necessarily clash with the commonsensical view of supererogation, namely with the idea that someone may or may not do morally excellent deeds. However, some tension is raised by the idea that “what people *ought*³⁰ to do is (...) to maximize the good” or by the idea that “it can never be right to prefer a worse state of affairs to a better”.

Obviously the passage from “one may always commit actions with the best outcome” to “one ought to commit actions with the best outcome” is the one creating problems for a supererogationist position. This is a fairly obvious point of conflict between *allowing* a morally good deed (according to a supererogationist frame) and *requiring* a morally good deed. However, the conflict runs deeper than that and takes, I believe, a more complicated shape. For once we add several classical consequentialist and utilitarian features to this basic, compelling idea (in whichever formulation one might prefer) the conceptual landscape becomes more complicated and presents more opportunities of conflict. Even here, where I have tried to match

²⁹ Portmore (2003, p. 304).

³⁰ My italics.

two basic ideas, the basic idea of supererogation and the core intuition of consequentialism, things are more complicated than a mere conflict of permission vs. obligation. For what supererogation allows is not exactly the same thing that consequentialism requires. Involved in the commonsensical idea of supererogation (as the two-parts has made clear) view is just a notion of “better” moral deed (this is what is allowed but not required), not necessarily a notion of “the best”. In addition, the commonsensical notion of supererogation neither excludes nor involves the idea that it is *the outcome* of the action that is evaluated as “better”. Usually, in discussing supererogation, it is the moral action that is evaluated as “better” without mentioning if it is the outcome of the action or something else³¹ that made it “better”. As said before, once the full theoretical apparatus of consequentialism and utilitarianism are in place, on top of this basic intuition, the opportunities for conflict multiply.

Therefore, in order to make somehow clearer the differences, the conflicts and the transformations that follow the reconciliation attempts between supererogation and theories of consequentialist inspiration, I will provide below a chart of main points of conflict and how they tend to be resolved, namely by modifying the basic supererogation concept or the consequentialist theory.

a) The impersonal scale measuring the good

First, one might notice that in characterizing the two sides of supererogation (being neither obligatory nor forbidden and being morally better than alternatives) the second part is left vague; we are told that the supererogatory action is “better in some

³¹ As it was suggested, it might have been the sacrifice of the agent that were taken into account, not necessarily the consequences of his action, when the verdict of the evaluation comes out as morally better.

sense” but one is left to fill in the details as one chooses. This is exactly where consequentialist and utilitarian theories supply promptly detailed theoretical constructions, so that “better” tends to acquire a quite technical sense.

Act-consequentialism is generally characterized as a certain sort of view about the relation between an act's rightness and its consequences. An act-consequentialist holds that states of affair (outcomes, consequences) can be objectively or impersonally ranked according to their goodness and that any given act is morally right or permissible if and only if its consequences are at least as good, according to the impersonal ranking, as those of any alternative act open to the agent--the doing of an act being itself included among its consequences. (Slote 1984a, p.139)

The impersonal ranking of states of affairs is therefore used as a scale in determining the rightness of a moral deed: this is something completely new for a commonsensical view of supererogation. Supererogation seen through consequentialist lenses will have to adopt at least some parts of this way of accounting for what it means for one deed to be “morally better” than others. In this sense, doing a supererogatory action is “better” than not doing it because its outcome would rank higher on the impartial/impersonal scale of good states of affairs.

I think that the concept of supererogation is not, at first sight, incompatible with such a procedure of impersonal/impartial measurement. However, once the point of view of the agent comes into discussion (and it has to be taken into account because the agent is being prescribed a highly demanding kind of moral behavior), the impersonal character of the scale will become problematic, and several authors³²

³² Scheffler (1982), Slote (1984), Portmore (2003).

will look for a way to make room for the agent's point of view when measuring the goodness of the outcome.

b) Required optimization or maximization

Second, supererogation, in its commonsensical understanding, lacks not only explanations about what “better” moral deeds might mean, but also the idea of an optimal or maximally good moral deed. On the consequentialist side, this idea is one of the hallmarks of the theory:

Among ethical theories, those that I call 'act-consequentialist' may be characterized roughly as follows. Such theories first specify some principle for ranking overall states of affairs from best to worst from an impersonal point of view. (...) After giving some principle for generating such rankings, act-consequentialists then require that each agent in all cases act in such a way as *to produce the highest-ranked state of affairs that he is in a position to produce*.³³ (Scheffler 1982, p.1)

However, he adds that all kinds of consequentialism share these two features:

"ranking overall states of affairs impersonally and the general idea that the best states of affairs are *somehow* to be promoted".³⁴

An act-utilitarian is, according to the prevalent conception, an act-consequentialist with a particular view about how states of affairs are to be impersonally ranked: roughly speaking, the goodness of states of affairs depends only on the well-being, happiness, satisfaction, utility, or desire-fulfillment of the individuals who exist in those states of affairs and one state of affairs is better than another just in case it contains a greater sum of individual utilities, or a greater overall balance of satisfaction over dissatisfaction. Thus act-consequentialism holds that a right act must be optimific, and the act-utilitarian, in addition, that optimizing always means maximizing the sum of individual well-being, desire fulfillment, etc. (Slote 1984a, p.139)

³³ My italics.

³⁴ Scheffler (1982, p.2).

While the idea of a maximum of good in itself does not present an incompatibility with supererogation, the whole point in establishing a maximum is that it becomes, by the lights of the utilitarian theory, *obligatory*. Obviously, this is one of the features of consequentialist and utilitarian theories that comes into conflict with supererogation. It is already a commonplace³⁵ to point out the incompatibility between the requirement of maximization and supererogation: if supererogation is mainly about doing more than required, then to require the maximum one can do leaves no place for going beyond what is required. As the two (supererogation and the maximization requirement) seem to exclude each other, one is forced to say either that supererogation does not actually exist (and to explain away the commonsensical moral intuition upholding its existence)³⁶ or to say that consequentialist and utilitarian theories should give up optimization or maximization, respectively. The route usually taken is to accept that supererogation represents a robust intuition of common sense morality and the consequentialist inspired theories should adapt in order to accommodate it. In this sense, authors who want to stick to the basic intuition of consequentialism while being able to accommodate supererogation,³⁷ propose giving up the maximization requirement and adopt instead a satisficing requirement, i.e. the view that what is morally required is to do something “good enough” by some adopted standard, not something that is best. If this proposal indeed reconciles the commonsensical notion of supererogation with consequentialism, will be discussed in the following sections.

³⁵ See Heyd (1982), Mellema (1991), New (1974), Portmore (2003), Scheffler (1982), Slote (1984).

³⁶ See New (1974) and Vessel (2010).

³⁷ Slote (1984), Hurka and Schubert (2012), Dreier (2004).

There are authors such as Zimmerman (1993) and Vessel (2010) who disagree with this widely held verdict of incompatibility between supererogation and maximization requirement. Their strategy will be to try to keep the maximization requirement while introducing other changes in the conceptual frame surrounding the problem of compatibility of the two.

To conclude, this is a point where supererogation imposes some serious modification upon the consequentialist frame, the modification usually taking the form of a rejection of the maximization (or optimization) requirement.

c) The threshold of supererogation as satisficing

Obviously missing from any consequentialist or utilitarian theory is the idea of a threshold of what is required, beyond which one may permissibly act in order to obtain even better outcomes. On the other side, supererogation is actually defined by the existence of such a threshold. As a consequence, consequentialist theories trying to accommodate supererogation will usually adopt a kind of threshold of what is required below the optimal or the maximum possible. The kind of threshold to be discussed is the one given by the concept of *satisficing*, where the outcome is not supposed to be the best possible one, but one that is “good enough” in the circumstances. This proposal will obviously create room for going “beyond what is required”. Nevertheless, it will also leave some room for doubt and debate regarding the success of capturing the commonsensical moral intuition of supererogation by this theoretical device.

d) The outcome of the action and the sacrifice of the agent as a cost

The sacrifice of the agent and the overall (presumably) good result of his deeds are not quantified or measured against each other according to the commonsensical view of supererogation. In a consequentialist or utilitarian frame naturally the outcome will be important and it will be important for it to be measurable/quantifiable. The impersonal maximization of the good is an idea usually aiming at something like the “greater good” of all. The problem with supererogatory action is that the pursuit of the greater good by the agent might come with a heavy cost for the agent. The two-part image of supererogation (that is, permission and superiority of the supererogatory action) does not mention the sacrifice of the agent as a condition for something to be considered supererogatory. It only says that one is permitted not to do it; it does not say why. However, often times the classical paradigmatic examples of supererogatory actions are saintly and heroic deeds, which tend to have a heavy cost for the agent and to bring very good outcomes (for the rest). Even if sacrifice is not always present with supererogatory action, it is enough to have some cases of supererogation that bring very good results with heavy cost (e.g. the soldier who saves many lives by sacrificing his own life) in order to generate objections to theories of utilitarian inspiration. For in a classical “crude” version of utilitarian theory such heavy sacrifices will be required in case they maximize the result. Therefore, this trait of utilitarianism has become a classical objection usually labeled as the *demandingness objection*: it is too much to require from an agent in certain cases to maximize the goodness of the outcome. Critics point out that this is the point where the concept of supererogation, properly integrated, would bring some relief from exaggerated utilitarian demands by making sacrifice and heavy cost for the

agent optional. The manner in which authors attempt to make this integration is by leaving the agent some latitude regarding the allocation of her time and effort: this is where the agent-centered prerogatives make room for agent-favoring permissions and agent-sacrificing permissions in Scheffler (1982), Slote (1984b) and Hurka and Schubert (2012).

These points from a) to d) are sensitive points where supererogation and consequentialist theories collide. Authors who will try to make supererogation sit comfortably with consequentialism and utilitarianism will modify either or both: they will try to impose exceptions from the impersonal scale measuring the goodness of the outcome, they will try to give up the maximization requirement and adopt satisficing instead (which will function also as a threshold for supererogation) and they will try to calculate what kind of sacrifice would be reasonable to be demanded from the agent. All these points of conflict and modification will come out in the way authors state what they each take to be “the” problem of supererogation.

2.2. The Problem of Supererogation in Utilitarian Setting

The discussion up to this point has shown how the commonsensical notion of supererogation would look under consequentialist lenses and how the attempt to accommodate supererogation sometimes modifies the structure of theories of consequentialist inspiration. In short, it was about supererogation and consequentialist theories modifying each other when cohabitation is sought.

It remains to clarify, however, some of the consequences of this transformation for the problem of supererogation. The problem 'how are we permitted not to do an excellent moral deed?' seems to receive a *prima facie* categorical negative answer from the usual consequentialist-utilitarian perspective, namely that we are not allowed, if by “excellent moral deed” one understands “the action with the best available outcome”.

On closer examination, however, this vague and general formulation of the problem of supererogation takes, in this context, a slightly different shape. Adding an impersonal scale of good outcomes and a notion of the best available outcome (that is, modifying the problem according to the points a), b) and c) from the previous section), the general question becomes '*how are we permitted not to do an action which has the maximizing/optimizing outcome impersonally/impartially considered?*' But this is not to say that the authors discussing supererogation in a consequentialist setting focus on this problem with this single question in mind.

First, the authors discussing how to acclimatize supererogation in a consequentialist environment often do not have supererogation as their primary focus or concern. From various theoretical concerns the problem of supererogation emerges as a manifold of particular problems, each with their own focus and own formulation (narrower than the general frame provided by the question above). For example, because of their different aims, Scheffler's approach to supererogation³⁸ will be different from Slote's;³⁹ Scheffler's concern may be formulated as 'How are

³⁸ *The Rejection of Consequentialism* (Scheffler 1982).

³⁹ “Satisficing Consequentialism - I” (Slote 1984a).

consequentialists permitted not to take into account the impersonal scale measuring the goodness of states of affairs?' while Slote's would be best expressed, I believe, as 'How are consequentialists permitted not to perform the action providing the optimal (best) outcome ?'.

Second, depending upon how each particular theory will further specify the meaning of some terms (for example “satisficing” as opposed to “maximizing”), the resulting question about supererogation will be slightly different. For example, one may notice that 'to be allowed not to maximize' is actually a permission to do less than the best, i.e. a permission to do something that is 'good enough'. Therefore the general formulation of the problem of supererogation in this context may very well be 'how are we allowed to satisfice (i.e. not to maximize)?'; however, taking into account that “to satisfice” might have several different meanings (as suggested in Hurka (1990) and Slote (2004)) the question itself might be posing different problems.

Even though the authors discussing supererogation might refer to different problems than the one in the general formulation (i.e. 'how are we permitted not to do an action which has the best outcome impersonally/impartially considered?'), the general formulation is still useful, I think, for orientation purposes, namely for distinguishing between various possible strategies of acclimatizing supererogation in a consequentialist/utilitarian setting. Two strategies become immediately visible from the general question: one can give up the requirement of “the best outcome” (i.e. of optimizing/maximizing) from the first part of the question, or one can concentrate on giving up the trait of impersonal scale of measuring the outcomes (by having an agent-centered approach rather than an impersonal one). However, there are more

subtle strategies, not immediately made visible by the question. Other authors have concentrated upon spelling out in more detail the scale(s) of value according to which an action is considered “better” or “best”.

Consequently, I will present these three strategies as listed:

- A) *Giving up the optimization/maximization requirement* and adopting a *satisficing* requirement instead.
- B) *Giving up the impersonal scale measuring the outcome* (and adopt a more agent-centered approach).
- C) *Introducing more than one set of values* for measuring “the best” outcome.

My discussion will follow this division. In the next section I will present authors upholding the satisficing solution (Slote 1984a, Dreier 2004) and in the following section authors who want to modify the consequentialist frame by giving up the impersonal character of the scale measuring how good is the outcome of an action (Scheffler 1982). Then I will present Zimmerman (1993) with a rather unusual attempt of mixing supererogation with maximization.

2.3. Satisficing as the Reconciliatory Solution between Supererogation and Consequentialism: Michael Slote

Usually, when supererogation is taken into account in consequentialist contexts, the main concern is not the problem of supererogation, but some larger theoretical point. In Slote's case, supererogation, seen as a carrier of powerful

intuitions of ordinary morality, serves as a test for the consequentialist position, a test meant to reveal if these powerful moral intuitions can be acclimatized in a consequentialist environment.

There are, famously, a number of moral verdicts upheld by common sense morality and denied by consequentialist inspired theories. Slote himself quotes Bernard Williams and Samuel Scheffler as prime examples of authors criticizing consequentialism for its disconnect with important moral intuitions:

Moreover, critics of optimizing consequentialism have recently tended to focus on one particular way in which such consequentialism implausibly offends against common-sense views of our obligations of beneficence. They have pointed out that (optimizing) act-consequentialism makes excessive demands on the moral individual by requiring that she abandon her deepest commitments and projects whenever these do not serve overall impersonally judged optimality. For example, it has been held by Samuel Scheffler (and others) that it is unfair or unreasonable to demand such sacrifice of moral agents, and by Bernard Williams (and others) that such requirements alienate individuals from their own deepest identities as given in the projects and commitments they hold most dear, thus constituting attacks on their integrity (integralness) as persons. (Slote 1984a, p.157)

Slote accepts that the traditional form of consequentialism (i.e. the one requiring an optimization of the overall outcome) cannot accommodate some of these commonsensical moral intuitions. It is clear that the common moral sense will judge the optimizing requirement as too demanding, especially in cases of great sacrifices. It might be optimal for one soldier to die in order to save all the others, but this, the common moral sense claims, cannot be demanded and the soldier failing to optimize the outcome in this way cannot be blamed. However, claims Slote, this does not mean that consequentialism as such should be abandoned, for one may abandon the

optimality requirement without abandoning the basic intuition of consequentialism, namely that there is a connection between the rightness/wrongness of a moral deed and its consequences. According to Slote, at least this one intuition of the common sense morality regarding supererogation, namely the intuition that one is not always required to act in a manner producing the best outcome, may be accommodated in a consequentialist setting.

He claims that act-consequentialism has been regarded as a “unitary moral conception”, namely the conception that “the rightness of an act depends on whether it produces the best consequences impersonally judged”,⁴⁰ but this claim hides two claims that are conceptually separable: first, that rightness of an act depends only on how good consequences are, and second, “that the rightness of an act depends on its having *the best*⁴¹ consequences (producible in the circumstances)”.⁴² Therefore one may uphold the first without endorsing the second, i.e. one may maintain that only the good consequences make an act right, but also that these good consequences need not be the best possible ones, but only “good enough” to make an act right:

Could not someone who held that rightness depended solely on how good an act's consequences were also want to hold that less than the best was sometimes good enough, hold, in other words, that an act might qualify as morally right through having good enough consequences, even though better consequences could have been produced in the circumstances? (Slote 1984a, p.140)

Therefore, Slote wants to advocate a new kind of consequentialism, one that does not have the optimizing requirement. His strategy is to argue that a 'satisficing

⁴⁰ Slote (1984, p.140).

⁴¹ My italics.

⁴² Slote (1984, p.140).

consequentialism' (rather than the 'optimizing consequentialism') would agree more with intuitions of common-sense morality of benevolence and would be, in this way, more plausible:

And since the plausibility of various forms of consequentialism partly depends on how far their implications diverge from the deliverance of ordinary moral intuition, this new form of consequentialism may turn out to have some distinctive advantages over traditional optimizing forms of consequentialism. (Slote 1984a, p.152)

Slote borrows the notion of “satisficing” from the literature of economics and his first examples of an agent plausibly settling for less than the best, but nevertheless “good enough”, is the example of someone selling their house: they might accept not the best price but something that is deemed a satisfactory price and is offered more promptly. The reason for accepting the lower price, claims Slote, is not an anxiety about not being able to sell or an indifference towards money, but simply being content with “good enough”. A second example is “the snacker”: the person who chooses not have an extra snack offered for free even though they know they would enjoy it and they do not fear any bad consequences. They simply decide they had enough to eat: Slote considers this to be a kind of moderation, one that is not a form of asceticism and which “it is difficult to see why it should count as irrational”.⁴³ By analogy,⁴⁴ he claims, we can make equal sense of cases of moral satisficing. For

⁴³ Slote (1984, p.145).

⁴⁴ Dreier (2004) thinks that the analogy between rational calculus of advantages and the moral domain does not work because rational satisficing does not make sense, but it makes sense for the moral domain. His criticism of Slote is that if someone does not take an extra snack, like in Slote's “snacker” example, then that person is still maximizing her/his preferences because they prefer not to snack. In order to satisfice one would have to prefer to snack while abstaining from eating the snack: “the Snacker’s utility for the proposition that he snack is higher than his utility for the proposition that he abstains just in case he prefers to snack. If he prefers to abstain, then his utility for abstaining is higher. Unless Slote means us to understand his Snacker to prefer to snack, even though he chooses to abstain when the snack is readily available, his Snacker is indeed choosing the

example, a hotel manager helping out a homeless family by giving them a spare room has done nothing wrong if she did not offer them the best room. She did less than the best, but nevertheless a commendable gesture of benevolence by offering something that was good enough in the circumstances. Another example used by Slote is the doctor who volunteers to go to a country in need of medical help. However, he is not required to go to the country which is most in need; no matter where he chooses to go (maybe following a personal interest), his gesture would be commendable even if the outcome is not the best possible, but only good enough. Slote thinks that in this way, a consequentialist theory can accommodate supererogation because it creates a kind of threshold of what is required, one that the agent may go beyond: it was permissible for both the doctor and the hotel manager to optimize, to do more than the required “good enough” and that would have been supererogatory:

One of the chief implausibilities of traditional (utilitarian) act-consequentialism has been its inability to accommodate moral supererogation. But a satisficing theory that allows less than the best to be morally permissible can treat it as supererogatory (and especially praiseworthy) for an agent to do more good than would be sufficient to insure the rightness of his actions. Thus, if the person with special interest in India sacrifices that interest in order to go somewhere else where he can do even more good, then he does better than (some plausible version of) satisficing act-consequentialism requires and acts supererogatorily. But optimizing act-

option with the higher utility. He is not merely satisficing, but maximizing.” (p.141). However, this is not what happens in the case of moral satisficing – a detailed discussion of his position will be in the next section. Remarkably, Dreier believes that rational satisficing presents the same kind of puzzle as the puzzle of supererogation, only that in the case of supererogation the puzzle can be solved: “The impossibility of supererogation looks to me to be the same as the impossibility of rational satisficing. If I am right that something’s being better from a certain point of view is the same thing as one’s having a reason to do it (or bring it about, etc.) from that same point of view, then the coherence of supererogation is the same as the coherence of satisficing. Satisficing appears to ignore reasons that one admits one has. And so does merely doing one’s duty when a supererogatory act was available. Yet satisficing theory says that it can be reasonable to satisfice, and commonsense morality says that it can be morally acceptable to fail to perform what is supererogatory. (p.148)

consequentialism will presumably not treat such action as supererogatory because of its (from a common-sense standpoint) inordinately strict requirements of benevolence. (Slote 1984a, p.157)

Slote admits that his solution needs elaboration. The “good enough” of satisficing needs specification. He reviews some conceptions about satisficing from Popper and Bentham and reaches the conclusion that a certain percentage below the maximum would probably be an appropriate conception of “good enough”.⁴⁵

The viability of Slote's solution is not the important point here, so I am not going to try to decide if his proposal of reconciliation really succeeds. From the perspective of the problem of supererogation, I believe there are two observations that are important.

First, the conflict he is trying to resolve is not about the demandingness of the optimizing consequentialist theory. The demandingness objection is a fairly common one against traditional utilitarian and consequentialist theories. The accusation against theories of this kind is that they ask too much from the agent by making the optimal or the maximal result obligatory because sometimes the best result comes with a very large cost for the agent (for example, paradigmatic cases of saints and heroes involve extreme sacrifices). And common sense morality objects that such large costs cannot be demanded. This objection will be discussed in detail in the next section, where the

⁴⁵ Hurka (1990) is critical of Slote here because, Hurka claims, Slote is equivocating between two possible meanings for “satisficing”: the “absolute” and the “comparative”. The absolute satisficing simply establishes a threshold of good enough action without reference to other alternatives present to the agent. The comparative satisficing demands that the outcome of the action be “reasonably close to the best” and so it makes reference to the maximum possible. Hurka believes only the absolute kind of satisficing would work for Slote's version of consequentialism because in a bad situation only absolute satisficing would make sure that enough has been required from the agent, enough to improve the situation significantly; the comparative satisficing, he claims, would only ask for some kind of improvement on the bad situation (which would be some percentage of the maximum).

impersonal aspect of consequentialism comes into play. Here it is important to notice that the objection discussed by Slote against optimizing consequentialism is not the same as the demandingness objection: the incongruity between common sense morality and optimizing consequentialism discussed here is simply that common sense morality accepts without problems that the agent might do less than the best (even when there is no significant sacrifice involved on the part of the agent):

So the divergence between common-sense morality and standard (utilitarian) act-consequentialism with regard to such cases cannot be accounted for in terms of a disagreement over whether one can correctly require an agent to sacrifice his own desires, projects and concerns in the name of overall optimality. (Slote 1984a, p.151)

Slote points out that his example of the hotel manager helping out the homeless family is an illustration of this idea: there is no major sacrifice in this case on the part of the hotel manager, but common sense morality and the traditional consequentialist view still diverge regarding the optimization requirement. Therefore, the agent may do less than the best not because the optimization requirement asks too much (much more than it is reasonable to ask) but simply because it seems perfectly acceptable and reasonable to do less than what optimization requires. In short, one is allowed to satisfice, not only when there is a large cost for the agent (and because of that) but simply because it appears to be a reasonable option.

Secondly, and more important, Slote points out that the concept of supererogation obtained by giving up the optimizing requirement *is not the same supererogation concept* featured in the commonsensical view about supererogation:

Such consequentialism in effect then allows various sorts of compromise between the demands of impersonal morality and personal desires and commitments. To that extent, it allows greater scope for personal preferences and projects than traditional optimizing act-consequentialism does. However, it offers less scope than would be available on most common-sense views of what an agent may permissibly do. For ordinary morality would presumably allow an agent (capable of doing better) to pursue projects that do not contribute very much to overall human well-being, and satisficing consequentialism - unless it maintains a very weak view about what it is to do enough good - will rule such projects out. (Slote 1984a, p.158)

The main point here, I think, is that even when one has made some room for going beyond what is required (i.e. when satisficing is required, optimizing is not), this might be a narrower scope for supererogation than the one attributed by the commonsensical view. The doctor volunteering to take care of patients in a country that she prefers, will behave supererogatorily if her plans change and she agrees to go to a different country, where help is most needed – this will be the verdict of satisficing consequentialism. However, the doctor *is required* to reach this satisficing threshold (that is, to go to a country in need), which means that satisficing consequentialism will *require* some acts of benevolence that might be declared by common sense morality to be entirely *optional*. There is also a secondary point, namely that this restricted concept of supererogation is a kind of compromise between the commonsensical notion of supererogation and the impossibility of supererogation (within optimizing consequentialism). This compromise also makes room not only for supererogation but also for a more personal approach to morality. To recall, the consequences of an action were to be measured on an impersonal/impartial scale of goodness for states of affairs. Slote is not arguing, like Scheffler (1982), that one may

ignore at times this scale (because of an “agent-centered prerogative”) but he is saying that given more choice (by not being under the obligation to optimize), agents may choose something closer to their own plans and aspiration when they are only under a satisficing obligation. The doctor in Slote's example may choose, in this way, a country she is interested in, instead of the country most in need of medical assistance.

In conclusion, I think Slote's answer to (his way of seeing) the problem of supererogation is rather complex. To recall, the generic problem of supererogation in a consequentialist context was formulated as '*how are we allowed not to do an action having the best outcome impersonally considered?*'. Slote's version of the problem of supererogation is rather something like '*how is a consequentialist allowed not to optimize (i.e. allowed to satisfice)?*' - which is the same as asking if a consequentialist is able to allow the existence of some sort of supererogation, because the obligation to optimize would hardly leave any room for supererogation.⁴⁶ One feature of his answer is that the consequentialist is allowed not to optimize because the main thesis of consequentialism (i.e. the connection between consequences and the rightness of an action) can be separated from the optimization requirement and one might want to drop the optimization requirement in order to be able to better match the intuitions of common sense morality; according to common sense morality there is no difficulty in accepting that someone who did less than the best in the circumstances has, nevertheless did the right thing and has done something that was “good enough”.

⁴⁶ There are authors who do not agree with the incompatibility between optimization/maximization requirement and the concept of supererogation; see Vessel (2010) and Zimmerman (1993).

Another facet of his answer is the rationality aspect: it seems that there is nothing irrational in economic thinking or in moral thinking about wanting less than the maximum possible. And therefore the answer to 'why are we allowed to satisfice?' seems to be simply “because there is nothing irrational about wanting less than the most one might be able to acquire”. This tenet might be more controversial, and Slote himself acknowledges the existence of a robust philosophical tradition connecting rationality with maximization of satisfaction:

Even those opposed to consequentialism and utilitarianism as moral theories have tended to think that (extramoral individualistic) rationality requires an individual to maximize his satisfactions or to do what is best for himself. (Slote 1984a, p.141)

He does not elaborate this point in his “Satisficing Consequentialism”, but he tries to make a case for the rationality of satisficing elsewhere, namely in “Two Views of Satisficing” (2004). He brings a virtue ethics perspective to this problem, his basic idea being that always looking to maximize might not necessarily bring more good or a better quality of life (because lacking the virtue of moderation does not always result in enjoyment).

2.4. Dreier's Moral Satisficing

An outline of Jamie Dreier's position is conveniently given in the title: in “Why Ethical Satisficing Makes Sense and Rational Satisficing Doesn't” he claims that while satisficing does not make sense for matters of prudential value and

economic context, satisficing makes sense in a moral context. This is an unusual position to have if one keeps in mind that the concept of satisficing was borrowed from economics literature. Also notably, he criticizes Slote for his assumption that examples taken from the domain of economics could be made analogous to the ones from the moral domain: Dreier claims that the “good enough” of someone selling a house on the spot for a lower price is not the same as the “good enough” of someone offering *not the best* room, but *a room* to a homeless family, even though they might present a *prima facie* similar structure. The similarity consists in the fact that both are examples of agents choosing an option that, they admit, weighs 'less' on the scale of good than another option available to them.

Briefly, the reason why he thinks rational satisficing does not work is that in normal, rational cases, agents will always maximize their preference, never satisfice. What Slote and others have described as cases of satisficing are actually, according to Dreier, cases of maximizing one's preferences. For let us take the example of the person selling their house. If they sell it for a lower price because they do not want to wait longer in incertitude, then this is their preferences, so they are actually maximizing their preferences even if they accept a lower price (because the utility is higher for a satisfied preference, even if the price obtained is lower). In order to truly satisfice, Dreier claims, the person selling the house would have to have a certain preference to which they attach the highest utility (e.g. to obtain a higher amount of money) and then to go against that very preference (i.e. to accept less). In short, in order to truly satisfice, one would have to prefer more money and then accept less money (which has an irrational air about it). Instead, what typically happens in

examples like this one, claims Dreier, is that one prefers time and so accepts less money, which is perfectly rational and this is why the example sounds plausible, but it is a *maximizing* example, not a *satisficing* one;⁴⁷ in order to satisfice one would have to accept less of the same thing that one prefers.

It is the same for Slote's example of the “snacker”: if the person does not take a snack then this is their preference and the higher utility is attached to it; in order to satisfice they would have to actually prefer the snack and at the same time abstain from it:

(...) the Snacker's utility for the proposition that he snack is higher than his utility for the proposition that he abstains just in case he prefers to snack. If he prefers to abstain, then his utility for abstaining is higher. Unless Slote means us to understand his Snacker to prefer to snack, even though he chooses to abstain when the snack is readily available, his Snacker is indeed choosing the option with the higher utility. He is not merely satisficing, but maximizing. Could the Snacker be choosing the option that he does not prefer, all things considered? It is hard to see how that is possible. If it is possible, then it is hard to see how it could be rational. (Dreier 2004, p.141)

Dreier maintains that rational satisficing and supererogation⁴⁸ are confronted with similar puzzles because they both describe kinds of behavior where the agent chooses to ignore certain reasons (if it is granted that “option x being better” is the same as “having a reason to choose x”) and that is difficult to explain:

The impossibility of supererogation looks to me to be the same as the impossibility of rational satisficing. If I am right that something's

⁴⁷ For how our intuitions might incline towards supporting a “maximizing” or “satisficing” verdict depending upon the kind of example and context we use, one can read Jenkins and Nolan's article “Maximizing, Satisficing and Context” (2010). Their thesis, in short, is that one will side with “satisficing” if “best” is understood in context as meaning “the few at the top” and with maximizing if “best” is understood as meaning “the one at the top”.

⁴⁸ He might imply that supererogation *is* a kind of satisficing.

being better from a certain point of view is the same thing as one's having a reason to do it (or bring it about, etc.) from that same point of view, then the coherence of supererogation is the same as the coherence of satisficing. Satisficing appears to ignore reasons that one admits one has. And so does merely doing one's duty when a supererogatory act was available. Yet satisficing theory says that it can be reasonable to satisfice, and commonsense morality says that it can be morally acceptable to fail to perform what is supererogatory. (Dreier 2004, p.148)

Supererogation, presumably, involves a kind of "better", the moral "better" choice that the agent is allowed to ignore. It is, therefore, a kind of moral satisficing where even if the agent accepts one choice is better than another, the agent can still not choose the better one. It would seem that it is the same structure as rational satisficing where one has a preference, i.e. something one considers as "better" but satisficing allows the agent to ignore the reasons making the choice "better" and settle for less.

The structure invoked here by Dreier is, I believe, one of conflict of reasons. For let us say that an agent may choose between A and B as courses of action and the agent admits that B is, in some sense, better than A; by hypothesis, the agent has reason to choose B. Then, if the agent chooses A, this is, at least at first sight, difficult to explain. Both rational satisficing and supererogation may be said to follow this pattern: the agent is allowed to refuse what she herself said is the better choice or chooses what she admits to be less good, ignoring reasons that were previously accepted as good reasons. It would seem that in deliberation the agent has decided that the overall balance of reasons inclines towards B and nevertheless is allowed to

choose A without any penalty or accusation of irrationality or inconsistency or (in the moral case) harm done:

Morality, we are inclined to think, is a matter of what reasons one has from the moral point of view. When there is a supererogatory act available, it would be better for you to perform it. So surely you have a reason, from the moral point of view, to perform the act. You may have some reason not to perform it, but at least typically you will have no reason from the moral point of view to refrain from it (if you do have some such reason, then it will ordinarily be outweighed by the reason you have to perform, because by hypothesis it is better to perform). But now it is hard to see how it could be permissible, from the moral point of view, to refrain from doing something that you have an undefeated reason (from that very point of view) to do. (...) In what sense is it “all right,” “permissible,” “not wrong” to fail to act? There seems to be no sense at all. (Dreier 2004, p.148)

However, according to Dreier, while supererogation may benefit from a reasonable explanation of this puzzle, rational satisficing may not. His explanation for why there is a way out of the puzzle in the moral case is that there can be at least two points of view in a moral situation, while there can be only one point of view in the rational satisficing, the view of “all things considered”. To explain the two points of view he presents the reader with the example of a stranger losing their hat by accident; should one try to help by fetching it back (which would be supererogatory) or not? According to one point of view (the beneficent point of view) one should try to help, while according to the justice point of view, there is nothing to be said in favor of fetching the hat:

Still, I want to suggest something like that explanation. Maybe there isn't just one “moral point of view.” Maybe there are (at least) two. To borrow from virtue theory, one point of view we can adopt is the point of view of the perfectly virtuous agent (...); or, less ambitiously, just the beneficent agent. From this perspective, there is everything to

be said in favor of fetching the stranger's hat and nothing to be said against it. Failing to fetch the hat is falling short of perfection (in this dimension) and not permissible at all. But we can also adopt a less ambitious perspective – that of the just person (maybe 'dutiful' would be a better word). From the point of view of justice, there isn't anything to be said in favor of going to all that trouble to get a stranger's hat. If I do go to all the trouble, that doesn't make me more just. The suggestion is that judgments of wrongness are made from the point of view of this less demanding virtue, whereas judgments of what would be better or worse are made from the more ambitious point of view. (Dreier 2004, p.149)

The two moral points of view are a solution to the puzzle because now there is an explanation why one would praise a course of action as 'better' only to desist from that course of action: option B is better from a point of view that is different from the point of view which favors A. In choosing A, one is not acting against their own accepted reasons, but one is simply choosing another reason, from another point of view. In short, one point of view recommends A and another point of view recommends B; B is "better" only from the second point of view and, therefore, there is nothing irrational in choosing A if one chooses another point of view. A possible objection could be that even so, the balance of reasons in general must have been inclined one way or the other and the permission to choose as one may please against it or in accordance with it, is still strange. That is, when balancing A against B the two points of view are confronted and one must win. To this Dreier replies that the balance of reasons, i.e. the overall or 'all things considered' point of view does not need to be a moral perspective. Therefore the two moral points of view may preserve their distinctive appeal without one being declared the winner in their confrontation:

In any case, my suggestion for how supererogation is possible is that there are two perspectives from which the two different sorts of

ethical standards are derived. What looked like a contradiction is merely a conflict: From the one perspective certain facts are reasons, and from the other they are not. Although it is true that at some point the two perspectives must be commensurated, the conflict is not reintroduced as a contradiction because the point of view from which we make our all-things-considered judgment needn't be an ethical perspective. (Dreier 2004, p.150)

Dreier thinks that rational satisficing cannot divide this way between points of view because the perspective of rational satisficing is simply the perspective of "all things considered".

I will not discuss if Dreier is right and moral satisficing is different from rational satisficing. The main point of this section is to see what are his formulation and solution to the puzzle of supererogation as identified by him. The problem of supererogation is not necessarily posed in a utilitarian context, as Dreier does not pay attention to outcomes of the action or how one is supposed to measure them. He just assumes that one is able by whatever means to ascertain that to act supererogatorily is in some sense better than not, and he leaves the problem at that. Therefore the specific utilitarian formulation, namely 'How are we allowed not to do an action with the best outcome impersonally evaluated?' is not relevant for his considerations. For even though he is preoccupied with maximizing and satisficing, his problem seems to be a meta-problem: it is not about the actual question if/how one is allowed to do less than the good available to an agent and the moral complications of this situation; rather, it is about the concept of 'satisficing' and if it makes sense. One might think that Dreier's specific question is 'How are we allowed to choose an action that is just morally satisficing, not maximizing?' However, Dreier does not worry about the moral merits

of satisficing *versus* maximizing or how is one supposed to establish a satisficing threshold, or about fitting supererogation in a consequentialist frame (like Slote). His main worry seems to be the threat of irrationality for an agent who admits B is morally 'better' (in some sense) than A and then is allowed to choose A.

Therefore, I think that the problem of supererogation in this case is something of the form 'How is it not irrational for an agent to admit that an option is 'better' than the other and to act against this evaluation by choosing the lesser option – in both moral and prudential domains?' Dreier's explanation for the moral domain is that there are at least two points of view in the morality domain and therefore, what appears as 'better' from one point of view does not have to be so from the other.

2.5. The Impersonal Trait of Consequentialist Evaluations

When confronting the question 'How are we permitted not to do an action with the best outcome impersonally evaluated?' some authors have chosen to drop the optimization requirement, namely that one should always choose the action with the best outcome – and the result was that satisficing emerged as a solution. To recall, the solution was that we are allowed not to optimize/maximize because less than the best could be 'good enough' from a moral point of view. Once the requirement is established at the 'good enough' level, the agent may be said to go beyond what is required and therefore some room is made for supererogation (even though it might not have the same wide scope as the commonsensical notion of supererogation; some

things required by satisficing might end up as purely optional from the point of view of common sense morality).

This is not the only strategy available in order to acclimatize supererogation in consequentialist and utilitarian environments. Another strategy, sometimes involving satisficing, is based on an objection against the *impersonal and impartial* kind of evaluation demanded by consequentialist and utilitarian frames. At first sight, it does not seem to be a connection between supererogation and the impersonal aspect of consequentialist/utilitarian evaluation of states of affairs. However, once the agent is allowed to depart in various ways (e.g. by being allowed to give more weight to his own preference) from the impersonal/impartial way of measuring the overall goodness of the outcome, the obligation to maximize is implicitly dropped and this usually makes room for a notion of supererogation.

The objection against impartiality has a tradition and is well known as coming from Bernard Williams. He describes impartiality in morality in general as being

(...) something which, indeed, some thinkers have been disposed to regard as the essence of morality itself: a principle of impartiality. Such a principle will claim that there can be no relevant difference from a moral point of view which consists just in the fact, not further explicable in general terms, that benefits or harms accrue to one person rather than to the other. (...) from the moral point of view, there is no comprehensible difference which consists just in my bringing about a certain outcome rather than someone else's producing it. (Williams 1981, p.96)

Impartiality is not a fault of utilitarians only; theories of Kantian and deontological persuasion have also thought that being impartial and impersonal are main ingredients of being moral. Nevertheless, utilitarianism has a specific way of

demanding impartiality, one implied by the way the sum of individual utilities is supposed to be calculated. Nobody's happiness is supposed to matter more and the only thing that matters is the maximization of the total, a total in which any kind of individuality is lost (as in any sum total). The acute problem of this view becomes evident when the utilitarian demands that any resource of time and energy be dedicated to the maximization of the sum total. It would seem that there is no room left for the agent to be involved in personal projects that do not bring sum total maximization of utility:

It is absurd to demand of such a man, when the sums come in from the utility network which the projects of others have in part determined, that he should just step aside from his own project and decision and acknowledge the decision which utilitarian calculation requires. It is to alienate him in a real sense from his actions and the source of his action in his own convictions. It is to make him into a channel between the input of everyone's projects, including his own, and an output of optimific decision; but this is to neglect the extent to which his actions and his decisions have to be seen as the actions and decisions which flow from the projects and attitudes with which he is most closely identified. It is thus, in the most literal sense, an attack on his integrity. (Williams 1981, p.117)

This kind of critique of utilitarianism, about the tendency of utilitarian theories to require the atrophy of the personal, has prompted attempts of reconciliation between the impersonal approach and an agent-centered approach. Such an attempt is Scheffler's agent-centered prerogative.

2.6. Scheffler's Agent-centered Prerogative and Making Room for Supererogation

Scheffler (1982) lists two classical objections to theories of consequentialist/utilitarian inspiration, one of the “view from nowhere” regarding the distribution relations between agents in a utilitarian scenario and the other from the point of view of the agent making decisions in a consequentialist manner. The first one is that utilitarianism prescribes ignoring the unhappiness of a few if this leads to a maximization of overall happiness. The second is Bernard Williams' worry about utilitarianism resulting in alienation from one's own life projects. Scheffler argues that Williams' worry about alienation is rather vague. If it means that utilitarianism demands that sometimes we give up our own plans when these would extract a large cost or impose too much damage on the others, then this does not seem wrong and it is something required by pretty much all non-egoistic moral theories, not only by utilitarianism. However, a charitable interpretation of Williams' objection would be, according to Scheffler, that alienation takes place, not because the agent is required to *give up to some* of his projects, but because the agent is required to evaluate *all* personal projects by reference to the impersonal scale measuring the capacity to increase the overall goodness or happiness; this is quite unnatural because this is not how we usually evaluate our projects:

Utilitarianism thus requires the agent to allocate energy and attention to the projects and people he cares most about *in strict proportion* to the value from an impersonal standpoint of his doing so, even though people typically acquire and care about their commitments quite independently of, and out of proportion to, the value that their having

and caring about them is assigned in an impersonal ranking of overall states of affairs (Scheffler 1982, p.9)

Notably, Scheffler considers this second objection as an objection not only against utilitarian theories, but also against any kind of consequentialist theory because they share the theoretical feature of impartial ranking of overall states of affairs.

Scheffler's proposal for tackling this difficulty is the introduction of an agent-centered prerogative which is meant to make it permissible for the agent to spend resources on projects evaluated *out of proportion* with the impersonal scheme:

On a plausible view of this kind the answer to the question of whether an agent was required to promote the best overall outcome in a given situation would depend on the amount of good he could thereby produce (or evil he could avert), and on the size of the sacrifice he would have to make in order to achieve the optimal outcome. More specifically, I believe that a plausible agent-centered prerogative would allow each agent to assign a certain proportionately greater weight to his own interest than to the interests of other people. It would then allow the agent to promote the non-optimal outcome of his choosing (...) (Scheffler 1982, p.20)

Not any kind of project may receive this permission, of ignoring the impersonal ranking, but there will be some. I will not enter into the details of his proposal here, like the circumstances in which it is permissible to ignore the impersonal ranking, how the greater weight for the agent's plans will be assigned, difficulties of this view and so on. Whatever the details of this proposal, one thing becomes clear, there is room made for supererogatory conduct:

Since it would permit people to devote energy and attention to their projects and commitments out of proportion to the weight from the impersonal standpoint of their doing so, the view would lack the

feature that generates that objection. But at the same time, it would certainly on such a view always be permissible for an agent to bring about the best available state of affairs. Thus there might be an agent who willingly sacrificed his own projects for the greater goods on this view his conduct would be supererogatory. (Scheffler 1982, p.22)

Scheffler main focus here is not, obviously, supererogation. Rather, the main concern appears to be finding a way to meet certain objections (related to the impersonal/impartial way of evaluating outcomes) raised by common sense morality against consequentialist and utilitarian theories. This is an enterprise similar to Slote's (1984), but while Slote was *keeping the impersonal ranking* of states of affairs and was proposing a lower threshold for what is morally required *on that scale*, Scheffler wants to proclaim the agent's *independence from the impersonal scale* in certain circumstances (circumstances in which the scale may be legitimately ignored).

Scheffler's main question might be said to be 'How could consequentialism allow the outcome of a moral action not to be evaluated according to the impersonal/impartial ranking of the states of affairs?' His solution, (that the agent has sometimes the prerogative to give more weight to his own plans and therefore is not under the obligation to perform the action with the best result impartially considered) becomes only tangentially also a solution for the problem of supererogation. However, it is, without doubt, one way in which supererogation may fit in an utilitarian frame, thereby providing a possible answer to the question 'How are we allowed not to do an action with the best outcome impartially/impersonally evaluated?'.

The answer is that the agent is allowed sometimes not to take into consideration the obligation of optimization/maximization because the agent is allowed sometimes to evaluate the outcome of his action independently from the evaluation on the impersonal scale. This is the agent's prerogative, to be morally permitted in certain circumstances to ignore the demand of producing the impersonal best ranking outcome and thereby to ignore the maximization demand. In circumstances where the agent has this prerogative, she might choose nevertheless to take into account the demands of the impersonal ranking at a cost for her own plans. This is the spot where supererogation falls into place: when the agent is allowed to choose between ignoring and taking into account the demanding impersonal scale, the agent might choose the "greater good" of the impersonal demands and act in this way supererogatorily. The examples are not difficult to find: Slote's doctor may choose to go to a country where her medical help is going to have the greatest impact (impersonally judged) or in a country where the impact is more modest, but the doctor might have a personal interest in being there; in this case, going to the first country, where the impact is greatest, is supererogatory.

There are several assumptions here that Scheffler does not discuss.⁴⁹ In a comparison of the outcomes of personal plans with outcomes of the actions prescribed by the impersonal ranking, the assumption seems to be that the impersonal ranking is going to be more demanding, i.e. it is going to ask for more things to be sacrificed. This needs not to be always the case. For example, a personal plan may sometimes involve much more sacrifices of well-being than the impersonal demands of morality:

⁴⁹ This is not a criticism, nor, for that matter, surprising since supererogation was not his main topic of discussion.

the impersonal evaluation of outcomes might require the doctor to go to the country most in need, but joining a religious order might require going to the county most in need, celibacy and asceticism. Another supposition added to the one that the impersonal ranking would be the most demanding is that sacrifice on the part of the agent is an indication of morally better outcomes impersonally judged. This needs not be always the case either.⁵⁰ The larger picture of presuppositions seems to be divided in two: on the one hand, actions made according to personal plans would presuppose less sacrifice from the agent and moral outcomes scoring lower on the impersonal scale of the good (i.e. the outcomes would be less good for others but better for the agent); on the other hand, actions made by accepting the demands of the impersonal scale would presuppose more (or more serious) sacrifices from the agent and they would score higher, even highest on the scale (i.e. the outcomes would be best for others but less good for the agent). This is a common assumption, but by no means an unassailable one. Especially in a consequentialist framing, one cannot take for granted that the greater the sacrifice of the agent, the greater the value of the outcome will be.

However, Scheffler's claim that there is room made for supererogation by his agent-centered prerogative is not affected by the issue I have just raised. What is new and remarkable about his way of seeing supererogation is that there is no threshold, properly speaking, that the agent would go above and beyond. The agent is presented with a choice (in the circumstances where the prerogative is allowed) of doing more for herself (i.e. personal plans) or more for others (paying heed to the impersonal ranking). She is permitted to do either and there is an option where she would do

⁵⁰ For an illuminating discussion of how sacrifice might lead to worse overall results, one might see Jean Hampton, "Selflessness and the Loss of the Self"(1993).

more by impersonal standards. But in choosing to take into account and obey the obligation presented by the impersonal scale, she does not go beyond an established threshold of obligation that was lower *on the same scale*. She simply has chosen *another scale* (a more demanding one). This is a concept of supererogation that is different from the commonsensical one. And not only is the concept of supererogation at issue here slightly different from the commonsensical one: the question is also slightly different. For Scheffler's agent-centered prerogative is not exactly an answer to the general question 'How are we allowed not to do an action with the best outcome impersonally evaluated?', but rather to the more specific question 'How are we allowed to ignore sometimes the demands of impersonal ranking of outcomes in moral action?'.

2.7. Zimmerman on Supererogation and Sets of Values

Zimmerman (1993) is one of the authors⁵¹ who believe that supererogation and the maximization requirement are not incompatible. Up to this point, the question 'How can we be allowed not to do an action with the best outcome impersonally evaluated?' has received various interpretations and various answers. All of them are more or less, attempts of making supererogation go along with consequentialist or utilitarian positions. Some authors assumed that the main problem was the requirement of 'best outcome', i.e. of *maximization* and they proposed *satisficing* as

⁵¹ One could also see Vessel (2010) "Supererogation for Utilitarians" for several varieties of utilitarianism accommodating various versions of supererogation.

an answer (like Slote (1984) and Dreier (2004)); other authors have assumed that the main problem was the requirement to have the outcome of the action impersonally evaluated and they proposed ways of making the impersonal measuring non-obligatory in certain circumstances (like Scheffler's prerogative); yet other authors have supposed that the main problem consists in making more room for permissions within the utilitarian frame (like Hurka and Schubert (2012) with agent-favouring and agent-sacrificing permissions) while giving a reason for the superiority of the supererogatory act.

Zimmerman's way of seeing the problem involves the difficulty of keeping the requirement of maximization while making room for (some concept of) supererogation:

The question of direct concern is this. Does a theory according to which what one ought to do is the best that one can do imply that supererogatory acts are impossible? (...) [I]t will be argued here that the implication does not hold. It will be argued, that is, that such a theory can in fact accommodate supererogation. (Zimmerman 1993, p.373)

This does not mean he does not acknowledge the difficulty of the problem and the likelihood of it being a challenging problem. The classical incompatibility between the two is well-known:

The point might be put this way. In those cases where it is left to the agent's discretion whether to do A or B, where A would be supererogatory and B not, A is superior in value to B. In such cases, maximizing theories imply that neither A or B is optional; on the contrary, A is obligatory (in the absence of any third alternative of equal or superior value) and B wrong. Thus maximizing theories imply that supererogatory acts are impossible. (Zimmerman 1993, p.375)

Zimmerman's reasoning starts from taking both maximization and the optionality involved in supererogation at face value: if one accepts both, then what would follow? If both A and B are permitted, the reasoning goes, then they must be equal in value, according to the maximization requirement (otherwise the maximization requirement would make one obligatory and the other wrong). But if they are equal in value, then the whole problem of supererogation actually consists in this: *how can both actions be equal in value (because both permitted) and still one of them be, in some sense, 'better' than the other?*

Somehow a maximizing theory must endeavor both to have its cake and to eat it – that is, endeavor both to say that A is superior in value to B and yet to say that A and B are of the same value. (Zimmerman 1993, p.375)

The only solution to this, Zimmerman claims, is to have different sets of values measuring the two actions. For without that, one would be forced to say that the two actions are both equal and unequal in value:

This can be achieved by, and only by, declaring that there is more than one set of values pertinent to the moral evaluation of an act. One set of values- call it the deontic set – must be said to be pertinent to the determination of right and wrong and obligation, the other – call it the non-deontic set – not. Where A is supererogatory but it is left to the agent's discretion whether to do B instead, a maximizing theory must declare A to be non-deontically superior to B but deontically equivalent to B. (Zimmerman 1993, p.375)

Zimmerman is proposing here only a hypothesis, namely that if supererogation is able to cohabit peacefully with the maximization requirement, then what would

make this cohabitation possible would be the existence of at least two separate sets of values measuring the value of supererogatory and non-supererogatory deeds:

If an act is to be good in a way that does not render it obligatory, its goodness must, it seems (but clearly more needs to be said here), reflect a value that its deontic status does not. (Zimmerman 1993, p.377)

To conclude, Zimmerman's way of solving the puzzle of supererogation in an utilitarian context is to assume that equally permitted actions have equal moral value and therefore we are obviously permitted to do one or the other. However, this answer leads to a further problem, the problem of explaining how between actions equal in deontic value, one could be superior in value to the other.

His take on the problem of supererogation, I think, would be best expressed by the question 'How can one explain that of two actions with the same value (because they are equally permitted) one is considered more valuable than the other?'. This is a quite different problem from the general problem of supererogation.

2.8. Conclusions about Utilitarians, Consequentialists and the Problem of Supererogation

The various strategies used to accommodate supererogation to consequentialism and utilitarianism are as many ways of understanding the problem

of supererogation in a different manner. Sometimes, a different way of understanding the problem brings along a different concept of supererogation, too.

When the point of conflict between consequentialism and supererogation is believed to be the maximizing/optimizing requirement, then, usually, satisficing is proposed as a solution that would make room for supererogation. The problem of supererogation, in this instance, becomes 'how can a consequentialist be permitted to satisfice (i.e. to not do the best)?' Even if one accepts that the justification given by Slote (1984) or Dreier (2004) is a reasonable one, the concept of supererogation allowed by satisficing has a narrower domain than the commonsensical one: some actions declared obligatory by the satisficing theory will still be merely permitted according to common sense. In Dreier's case, the answer requires a division of the moral domain in at least two sets of values, such that supererogation would make sense in a domain of beneficence, but not in one of justice. This tenet has already added at least one extra determination to the rather general commonsensical concept of supererogation: that it belongs to beneficence, but not to justice.

When the point of conflict between consequentialism and supererogation is believed to be the requirement to measure any outcome of an action according to *an impersonal scale*, then the solution is to allow for a more personal approach in measuring the outcomes and, therefore, to allow that in some circumstances agents are justified in not measuring their actions and plans according to the impersonal scale. In this way an agent-centered prerogative is created (Scheffler 1982). The problem of supererogation, in this instance, becomes 'how can a consequentialist be

permitted sometimes to ignore the impersonal scale and not do the best action available?

When the point of conflict between utilitarianism and supererogation is believed to be, not maximization, but the uniqueness of the value scale, then the proposal is to have at least two moral sets of values (Zimmerman 1993). One of them should justify why doing something supererogatory and not doing it has the same value and the other should justify why doing something supererogatory should be seen as “better”. In this case, the problem of supererogation becomes "How can one explain the possibility that of two actions with the same value (because they are equally permitted) one is considered more valuable than the other?"

2.9. Hurka and Schubert on Extending Permissions

Hurka and Schubert (2012) develop a theory that is not utilitarian or consequentialist, but is similar to the ones presented in this chapter because it aims at integrating commonsensical moral intuitions about the agent's autonomy into a theoretical framework. I believe their position is interesting because their project is conceived in the same spirit (of giving the agent more latitude in moral decisions) as Slote's or Scheffler's, but the theoretical means are quite different: they choose a frame of *prima facie* duties and permissions (derived from W.D. Ross). They also criticize Scheffler's solution for not being close enough to the commonsensical intuition about what is permitted to a moral agent, even if their thesis is quite close to

Scheffler's idea that the demands of the impersonal scale for measuring happiness can be sometimes ignored in order to accommodate commonsensical moral intuitions.

They identify two kinds of agent-centered permissions: *agent-favouring* permissions and *agent-sacrificing* permissions; each corresponds to an intuition of common sense morality. Agent-favoring permissions allow one to favor oneself within certain limits, even if this results in not obtaining the greater good:

(...) while you do nothing wrong if you give the five units to the other, you also do nothing wrong if you give the one unit to yourself. That's because you're permitted, at least up to a point, to prefer your own lesser to another's greater good, for example, your own lesser to her greater happiness. This agent-favouring permission underlies the common-sense idea that some acts are supererogatory, or beyond the call of duty. It would certainly be commendable of you to give the five units to the other person, but it's supererogatory rather than something you're morally required to do. And the reason it's supererogatory is that you're permitted to care somewhat more about yourself. (Hurka and Schubert 2012, p.2)

Additionally, they think that a second permission should be allowed: agent-sacrificing permissions, or permissions to care less about yourself.”⁵² It is also a permission *to not do the action that would produce best results overall*, but this time, when it is disadvantageous for the agent. For example, if the agent can give either two units of happiness to herself or one unit to another, then it is permitted for her to give the one unit to another even though the impersonally measured best result is not attained in this way. The authors are simply postulating the existence of such permissions and they take them to be underivative:

⁵² Hurka and Schubert (2012, p.2).

This account generates agent-favouring permissions by positing an independent and underivative permission to pursue your own good. More specifically, it holds alongside a prima facie duty, or duty other things equal, to promote everyone's happiness impartially, you have a prima facie permission to promote your own happiness. This permission has to be weighted against the impartial duty and in some cases will lose to it. (Hurka and Schubert 2012, p.5)

I will not discuss how justified this postulation is. The authors argue at length about the justifiability and the advantages of such a postulation. Supererogation, in this context, is seen as test, as one more proof that this is the kind of postulation needed in order to be able to include and justify supererogation within the theory:

We therefore see no sound objections to positing an underivative permission to pursue your own good, and an account that does so may have the further advantage of allowing a more complete account of supererogation. (Hurka and Schubert 2012, p. 8)

The authors claim that there are two main sides distinguishable in the conceptual realm of supererogation:

The concept of supererogation has two sides. On one side, a supererogatory act isn't morally required; on the other side, it's somehow better than its alternative, or "beyond" duty in a sense that connotes superiority. A complete account of the concept must capture this second side, explaining how supererogatory acts are better even though not strictly your duty. (Hurka and Schubert 2012, p. 8)

Therefore, the main task of a theorist who wants to account for supererogation is to explain „how supererogatory acts are better even though not strictly your duty.”⁵³ They claim that the “better” of supererogation cannot be explained quantitatively, by “more units of happiness”: supererogation is not better because it brings a better

⁵³ Hurka and Schubert (2012, p.8).

outcome in terms of happiness units; the “better” of supererogation has to be deontic in nature, as supererogation is a deontic concept.

The reasons behind this affirmation are unclear, but granting this, the authors come up with an original explanation of why supererogatory action is better than the alternatives. From their point of view, supererogation is a confrontation between *prima facie* permissions (the agent is permitted to favor herself) and *prima facie* duties (the requirements to promote other's happiness). Supposing that there are two choices available to the agent, A and B where B is the supererogatory one, the authors say that what explain B's superiority is the fact that B is not only a *prima facie* permission (like A) but *also a prima facie duty*. We end up having a choice between A and B because the *prima facie* duty to do B is defeated (when the *prima facie* permission is stronger than the duty), but this does not mean that it simply disappears; instead, it leaves the “trace” of a defeated *prima facie* duty:

Whichever choice you make you exercise a *prima facie* permission, but if you give the five units to the other you also fulfill a *prima facie* duty, and this may explain that act's superiority. It is deontically higher ranked because while equally permitted they are also *prima facie* required. A key feature of *prima facie* duties is that they don't disappear when they're outweighed. They remain as part of the moral situation and can leave what Robert Nozick calls “moral traces”. (...) A third trace, we are now suggesting, may be to explain why supererogatory acts are superior: though no more permissible than their alternatives, they fulfill a *prima facie* duty that the alternatives don't and are therefore on balance better. The *prima facie* permission opposing them makes them not required, but the *prima facie* duty favouring them makes them deontically higher-ranked.” (Hurka and Schubert 2012, p.9)

Therefore, while both A and B are permissible, B is better because it contains the trace of a duty: the fulfillment of a *prima facie* duty. While the authors thus offer their own solution to the problem of supererogation, they also admit that their account is not without problems. The difficulty comes from the second kind of permission, the agent-sacrificing permission.

But consider the case where you can produce either two units of happiness for yourself or one for another person. Here each of your option is unsupported by a *prima facie* permission, but if there is a *prima facie* duty to promote happiness impartially, giving yourself the two units also fulfills that duty and should therefore be preferable in the same way that supererogation is: it has more *prima facie* factors on its side. But this does not seem true: preferring your own greater happiness in this case doesn't seem better. The alternative of preferring the other's lesser happiness also does not seem better; your two options seem to be on a par. So our account of supererogation seems to have a counter-intuitive implication when extended to agent-sacrifice. (Hurka and Schubert 2012, p.17)

If the same theory of the “trace” left by the defeated *prima facie* duty is applied in this case, then the agent could equally choose between giving herself two units or giving someone else one unit, but giving herself the two units should be “better” because it fulfills a *prima facie* duty. However, the authors notice that this is a rather counter-intuitive result. For them, the two choices seem equally good: preferring your own greater happiness does not seem to be much better or much worse than preferring other's lesser happiness.

The authors resign themselves to this result, admitting that they do not have a solution for this conundrum.

While I do not claim to have a way out of this predicament, I might have an explanation. Sometimes, the duty to promote other's happiness, exactly because it is impersonal, is not sensitive to the difference between self and other. My suspicion is that common sense morality considers the distinction *actions benefiting the self* / *actions benefiting others* much more important morally than the division considered crucial here, namely between what promotes the greater good and what does not. One such observation was also made by Slote when he was speaking about the self-other asymmetry, namely the idea that what one may do for oneself is seen differently by the common moral intuition than what one may do for others.

For let us see the situation of agent-favouring permission and agent-sacrificing permission, side by side (with the examples used by Hurka and Schubert):

– *agent-favouring permission* – one has to choose either a) or b)

a) to self: 1 happiness unit – permitted under agent-favouring permission

b) to other: 5 happiness units – would lead by hypothesis to an impersonal better result, and therefore it is a *prima facie* duty and a *prima facie* permission.

– *agent-sacrificing permission* – one has to choose either a') or b')

a') to self: 2 happiness units – would lead by hypothesis to an impersonal better result, and therefore it is a *prima facie* duty and a *prima facie* permission.

b') to other: 1 happiness unit – permitted under agent-sacrificing permission

The authors rightly conclude that their result does not align with common moral intuitions: there is something strange about saying that giving yourself 2 units

of happiness is morally better than giving the other 1 unit. The authors also think that the two choices are not one better than the other, but they are morally equally good, i.e. “on a par”. I think that common sense, while disapproving their result, might give at least two alternative results: that the choices are equally good or that it would be morally better for the agent to lose 2 units of happiness in order to give 1 unit to another. No matter which one would be adopted, I believe that the discrepancy between the verdicts of common sense morality and the authors’ verdict is due to very different presuppositions about what is merely permitted and at the same time admirable or meritorious (i.e. paradigm cases of supererogation). I think it is predictable that the authors’ moral verdicts will not match the verdicts of common sense morality, given that the presuppositions of the two are so different. Let me explain.

To recall, meritoriousness or “being morally better” was established by the authors as the action resulting into the greater good: this is how the *prima facie* duties were established. The results, being measured according to an *impersonal* scale, are obviously *not sensitive* to a change regarding *the person* of the beneficiary of the action (that is, if it is the *self* or *the other*). On the contrary, common sense morality usually *is* sensitive to a change in person benefiting from the action, meaning that a change in person might bring a change in verdict: it is usually quite different if *I* am benefited by my action or *someone else* is; in both cases the action is allowed but the second one is far more likely to be meritorious or morally better. For example, common sense morality is more likely to label the first as selfish and the second as

altruistic. Therefore, the moral connotation of the action has changed depending on the person benefited by the action (self or other).

In short, my diagnosis of the conundrum is that the authors do not take into consideration what Michael Slote (1984b) has called “the self-other asymmetry”;⁵⁴ common sense morality, on the other hand, obviously follows the precepts of this asymmetry and gives its moral verdicts accordingly.

The asymmetry noticed by Slote seems to be nothing other than Hurka and Schubert's agent-sacrificing permission:

(...) if I have a choice between conferring a great benefit on myself or a lesser benefit on someone else, and these are the only relevant factors in the situation, common-sense morality tells us that it is quite permissible to sacrifice one's own greater benefit to the lesser benefit of another. In the absence of some special relation or obligation to that other, common sense might concede that it was irrational, stupid, or gratuitous to do so, but surely not that it was morally wrong. (Slote 1984 b, p.180)

However, if the asymmetry is followed closely, one might notice that the difference between what the *agent does to herself/what the agent does to another* becomes more vivid in a formulation about doing harm rather than in a formulation about benefiting. The agent may be equally allowed sometimes to benefit herself (this was the whole sense of Scheffler's agent-centered prerogative and the authors' agent-favouring permission) or benefit another, but this formulation of “benefits” obscures the losing side of this action. It might be the case that the agent is equally allowed to harm herself or to harm another (at least by omission) sometimes; let us say, by taking care

⁵⁴ Slote (1984b, p.181): “Thus ordinary moral thinking seems to involve an asymmetry regarding what an agent is permitted to do to himself and what he is permitted to do to others.”

of her own plans, the doctor does not go to treat patients in another country where help is badly needed. But the contrast is much sharper now between what harm I am allowed to do to myself and what harm I am allowed to inflict on others, because the range of the first will be much wider than the range of the second. Slote seems to agree with this, at least implicitly:

There is no fundamental moral reason why someone should not sacrifice *himself* to save five people who need organ transplants, and the side constraints built into ordinary morality concern only harm *done to others* in the name of good results. (Slote 1984b, p.183)

But what is more important in the self-other asymmetry is, I believe, that moral merit or excellency is also asymmetrically distributed, according to common sense morality: commonsensical morality will accept that sometimes the agent is permitted to care more about her own happiness but that the agent is equally permitted to care about other people's happiness or plans. But the fact that both actions are permitted does not mean that they will be equally praised: caring about oneself is not so morally admirable as caring about others. The difference is more dramatic in case of sacrifices. Sometimes the agent is equally permitted to lose or to impose the loss on someone else, but this does not mean that the moral merit is the same for the two actions. For common sense morality it is one thing to lose your own life to save five people but it is quite another thing to kill one person to save five. Even if both can be justified by the circumstances, the distribution of moral merit is widely different. It is important to notice that from the point of view of the impersonal utilitarian scale, the outcome looks quite the same for the two situations.

Analogously, in the authors' two cases (1 unit of happiness for self vs. 5 for other and 2 for self vs. 1 unit for other) the impersonal scale designates it as meritorious in the first case to benefit others and in the second case to benefit myself. From the point of view of the impersonal scale there is not much difference between the two as they are both meritorious because they have the merit of promoting the overall good, *no matter who the beneficiary is*. However, I believe that common sense morality, while not being a very consolidated, clear or coherent body of moral judgments, tends to be sensitive to who the beneficiary is because it tends to be rather other-oriented. Therefore morally excellent action tends to be a) something that primarily benefits another (even if it may benefit the agent, too); b) something that involves a sacrifice on the part of the agent (it is not necessary for an action to involve sacrifice to be morally excellent but usually, if it involves sacrifice benefiting the other, it is considered a morally excellent action).

One might notice that in the first case, i.e. the agent-favouring permission (one unit for self vs. 5 units for other), because it was about the agent being allowed to favor herself, another fact was obscured: that the morally excellent action (the supererogatory) involved a (small) sacrifice on the part of the agent (losing one unit of happiness). The commonsensical moral intuitions were running along with the authors' verdict in this case because the excellent action was a) other-oriented and b) involved self-sacrifice. Once these two traits change in the next case, common sense morality has a different verdict from the authors and the two ways of thinking become visibly at odds. In the second case, i.e. the agent-sacrificing permission (2 units for self vs. 1 unit for other), the intuitions of common sense do not correspond to the

utilitarian calculus any more because what the authors think of as an excellent action (giving two units of happiness to yourself and nothing to the other) is a) self-oriented, (not other-oriented) and b) does not involve self-sacrifice but a loss imposed upon the other. These two traits obviously go against what is usually appreciated in common sense morality: self-sacrifice and being other-oriented, namely against the self-other asymmetry concerning harm done – it is usually admirable to sacrifice yourself and it is usually blameworthy to impose the same kind of loss on others). These intuitions of common sense morality are the reason why I thought that a second verdict might be possible: in certain circumstances common sense might decide that it is better (more admirable) that the self loses 2 units of happiness in order to give 1 unit to another.

This explanation is, of course, not not an unconditional endorsement of the intuitions behind common sense morality, but an explanation for why the authors reach an impasse.

The discrepancy between the authors' moral evaluation and the common sense moral evaluation is, I believe, relevant for establishing the contours of the supererogation problem for Hurka and Schubert. They start with a classical utilitarian approach and try to modify this approach in order to accommodate commonsensical intuitions about what is morally permitted. Therefore the problem, in their own words is how can one explain “permissions not to do what will have the best outcome impartially considered.”⁵⁵ One might think that this is simply *their view of the problem of supererogation*, but things are a bit more complicated than that. Their answer to the main problem stated above is that agent-favouring permissions and

⁵⁵ Hurka and Schubert (2012, p.1).

agent-sacrificing permissions are basic starting points with numerous advantages (they are simply posited as underivative). One of the advantages of this postulation of permissions is that they can offer a solution to the problem of supererogation; therefore, the problem of supererogation is, one is lead to believe, something slightly different. Now, the problem of supererogation for Hurka and Schubert is not centered upon the question 'how and why do we have these *permissions* not to do what will have the best outcome impartially considered" because there can be no question about *permissions*: these permissions are simply posited, by hypothesis. The main question is, therefore, once we have these permissions, *how can we explain that one permitted option is better than the other* – that is the problem of supererogation in their view. In short, taking into consideration the two aspects of supererogation, the permission to do or not to do and the idea that it is better if one does act supererogatorily, the puzzle comes not from the permission aspect, but from the „better" aspect.

Their answer to this problem of supererogation is, to remind you, the “traces” left by defeated *prima facie* duties – that is what confers superiority on actions deemed supererogatory.

However, the difficulty encountered by this answer when confronted with intuitions of common sense makes evident another differentiation: not only the weight of the problem shifts, but the concept of supererogation obtained seems to be *different* from the commonsensical one. Common sense morality, might prefer as marks of supererogation the presence of self-sacrifice and being other-oriented. As discussed

above, these two ways of demarcating supererogation might coincide sometimes, but there are also cases where they fall apart.

Chapter Three: Reasons and Supererogation

3.0. Introduction

The previous chapter presented the consequentialist and utilitarian point of view on supererogation and its problems. The problems raised by supererogation in a consequentialist/utilitarian setting were focused especially on the requirement to optimize/maximize the good. I have discussed how the commonsensical way of understanding supererogation changed in several important respects.

First, as maximization was required, there was no threshold of duty to go beyond (as utilitarianism is especially prone to affirming the good-ought tie-up).

Second, there was little room left to talk about the agent's sacrifice in acting supererogatorily. The sacrifice entered in the calculus of utilities but it was seen as one loss compensated by the gain for many so it rarely mattered in establishing moral requirements (not to mention the situation where, if a sacrifice was made that did not bring maximization, its meritoriousness was usually lost for the utilitarian calculus). Slote (1984) and Scheffler (1982) are the ones trying to introduce theoretical devices

that would take into account the loss suffered by the agent when acting for the greater good.

The problem of supererogation, in these circumstances, was mainly a problem about how to explain/justify satisficing (i.e. abandoning the maximization requirement and adopting the “good enough” as required).

Commonsensical supererogation has a few classical traits: the moral excellence of the deed, the threshold of duty that one has to go “beyond” and (not necessarily, but typically) a sacrifice on the part of the agent. All these will emerge as different when embedded in the theoretical landscape of reasons. Let me try to briefly explain why – what changes are involved in speaking in terms of reasons.

The first chapter presented anti-supererogationist positions based on a Kantian idea of duty. Then, the second chapter analyzed utilitarian and consequentialist position about supererogation. These are the two classical ethical traditions that speak in terms of what is morally right, duty and obligation. When asked 'what one ought to (morally) do', the classical answers are “do your duty” or 'maximize the good (utility)'. Let us suppose, skipping over various difficulties, that an agent is in the happy case where she has already identified what it would mean for her to do her duty and to maximize utility: to keep a promise⁵⁶ and, respectively, to increase someone's well-being. The problem is that, in complex real situations, agents often times consider that they should do *both*, and these things come into conflict. Let us suppose that keeping a promise would bring a decrease in the well being of someone (or of

⁵⁶ I am following closely here an argument made by W.D.Ross (1930). His aim was to arrive at the concept of *prima facie* duty. I am using Ross because I think his position is an interesting intermediary between speaking in terms of duty and speaking in terms of reasons.

many) whereas not keeping the promise would at least keep well-being at the same level.

The whole point of the example is to say that, most of the time, the agent has complicated, competing obligations. When this happens, it seems reasonable and legitimate to have nuanced responses of the kind 'I will keep my promise if the consequences for others are not too harsh' or, alternatively, 'I am sorry but I will not keep my promise because the consequences are too harmful'. These ways of reasoning about (and responding to) moral common situations may suggest that classical ethical terms, like “duty” may be inadequate or inefficient in rendering these thoughts. W.D. Ross himself (whom I am following closely in giving this argument) changed the classical meaning of the term “duty” when he proposed his famous “*prima facie* duties”.⁵⁷ The alternative name given by Ross to *prima facie* duties is “conditional duty”, which clearly departs from the Kantian orthodoxy of duties being categorical.

The conditional form is not the only departure from the classical concept of duty, a departure obviously present in the kind of common moral reasoning like the one exemplified above. Usually, duties are understood as a matter of all or nothing: one can either do their duty or fail, there is no middle ground or degree. This classical picture leaves no room for a confrontation of duties, or for a complicated situation where, even though one did not fulfill their promise they omitted the act on moral

⁵⁷ The exact meaning of a *prima facie* duty is controversial, authors proposing two main interpretations: one where the *prima facie* duty is a tendency to make an action a duty, the other where a *prima facie* duty is something that would be a duty if nothing else would intervene. Ross (1930) gives the following definition: “I suggest 'prima facie duty' or 'conditional duty' as a brief way of referring to the characteristic (quite distinct from that of being a duty proper) which an act has, in virtue of being of a certain kind (e.g. the keeping of a promise), of being an act which would be a duty proper if it were not at the same time of another kind which is morally significant.” (p.19)

grounds and even so they feel sorry about not keeping their promise. This is a situation where, even though it seems that one has not fulfilled one's promise and has followed another (rival) 'duty', the duty of keeping a promise still holds some power over the agent's action because the agent feels *regret*. Disobeying other duties might carry even more weight by changing, not the action, but the *manner* in which the agent chooses to act.⁵⁸ These are subtle influences of 'duties' that seem to still have relevance for the action even when disobeyed. These nuances and the idea of weighing among competing duties in order to be able to have a resolution between rival claims, are difficult to express (to say the least) by using a classical notion of duty.

The utilitarian frame is no less rigid in this respect: in the classical frame there is no room for exceptions, extraordinary circumstances or failing to maximize the good *on moral grounds*. Regarding the utilitarian frame, Ross has the following illuminating critique:

In fact the theory of 'ideal utilitarianism' (...) seems to simplify unduly our relations to our fellows. It says, in effect, that the only morally significant relation in which my neighbours stand to me is that of being possible beneficiaries by my action. They do stand in this relation to me, and this relation is morally significant. But they may also stand to me in the relation of promisee to promiser, of creditor to debtor, of wife to husband, of child to parent, of friend to friend, of fellow countryman to fellow countryman, and the like; and each of these relations is the foundation of a *prima facie* duty, which is more or less incumbent on me according to the circumstances of the case. (W. D. Ross 1930, p.19)

⁵⁸ For a discussion of examples see Dancy (1993), Chapter 7.

As Ross makes clear, a *prima facie* duty may be more or less incumbent on the agent depending on the circumstances, which is another trait that does not fit into a picture of classical duty and obligation.

To summarize, the complicated moral picture that one might want to be able to describe involves, up to this point, the following: moral obligations that may be in conflict, obligations that may influence each other, obligations that may still influence the action after not being followed, obligations which have their strength dependent on circumstances, obligations that oblige only partially or that have a mere tendency to guide action.⁵⁹ These are all situations where *reasons* could work⁶⁰ as a more appropriate theoretical tool.

I think the following points may summarize the changes involved in employing reasons in thinking about moral obligation and moral discourse:

- a) The full force of categorical duty is dispersed in less categorical deontic forces that can come into conflict with each other or influence each other. Reasons are deontic entities, that guide action or say what one ought to do. But the deontic force of one reason is not the full force of duty in the classical sense. The 'ought' carried by a reason can be weakened, or enhanced, or overridden by other reasons.

⁵⁹ Again, I am following Ross (1930, p.28): "Any act that we do contains various elements in virtue of which it falls under various categories. In virtue of being the breaking of a promise, for instance, it tends to be wrong; in virtue of being an instance of relieving distress it tends to be right. Tendency to be one's duty may be called a parti-resultant attribute, i.e. one which belongs to an act in virtue of some new component in its nature. Being one's duty is a toti-resultant attribute, one which belongs to an act in virtue of its whole nature and of nothing less than this."

⁶⁰ Of course, reasons are employed in non-moral situations, too. But I will refer mainly to moral situations.

- b) A reason seen in this way, is a “partial” “ought” or something one ought to follow if there is no other reason against it.
- c) The answer to the question 'what ought one to do (morally speaking)?' is that the agent should follow the so-called balance of reasons, namely the result of comparing the weights of various reasons relevant to a given moral situation.

3.1. Reasons and Supererogation

To recall, the usual concept of supererogation has several constant traits: the moral excellence of the deed, the threshold of duty that one has to go “beyond” and (not necessarily, but typically) the sacrifice on the part of the agent. Each of these will suffer modifications within the frame of reasons.

A) The morally excellent and better action

Supererogatory action is action that is, at least in some respects, morally better: it is a morally excellent action. This creates a first problem: the good is a value, and value is usually thought to create reasons for action. Raz (1975a), for example, states explicitly that one of his presuppositions is that if some action *x* is good, then there are reasons to do it. The same presupposition is taken to be common sense by Dreier (2004). This is a problem for supererogatory action because reasons are also thought to be deontic entities, and therefore, from having a reason to do *x*, it also results that one *ought to do x* (if there are no other reasons against it). This seems

to be a classical case of good-ought tie-up, a frame where supererogatory action seems impossible (because most good action is not obligatory). There are several possible ways out (to be discussed in the next sections). One possible way out is to say that not all reasons are deontic entities (and therefore having a reason to x does not always imply that one ought to x). Another possible way out is to say that, in supererogatory cases, the sacrifice of the agent sometimes counts as a valid reason against the obligation to do x (Dancy 1993).

B) Going beyond duty and the frame of reasons

Obviously, one cannot speak in this context, about a threshold of duty beyond which actions are not obligatory anymore. Usually it is thought that an agent ought to do what the balance of reasons recommends. This is not, however, a mere substitution of roles: one could speak about going beyond duty, but it does not make much sense to speak about going beyond the balance of reasons. Most likely this happens because the balance of reasons is meant to indicate what is appropriate to do in circumstance. If a course of action is fitting, then doing 'more' than that seems nonsensical. I will discuss in the next sections Nagel's example (1986), in which the balance of impersonal reasons indicates the maximum that the agent can morally accomplish. Supererogation is possible only because this maximum is not required and the agent is allowed to go against the balance of reasons.

C) The Agent's Sacrifice

The theoretical framework of reasons, through its flexibility, allows for a proper theoretical treatment of the agent's sacrifice. In most cases discussed until now

the sacrifice of the agent in supererogatory action was taken into account as a *prudential* (i.e. not moral) worry that the agent might have. The frame of reasons allows for opposing, conflicting, lesser obligations such that the conflict is not necessarily seen to be one between duty and inclinations (or the greater good *versus* personal interest) – even if a version of this opposition is discussed in Nagel (1986). The agent may have agent- relative reasons, which can pull their own weight in a conflict of reasons. There is, therefore, a possibility that the personal reasons of the agent be endowed with deontic force, i.e. with moral dignity.

To see exactly how each of these traits may change the configuration of the problem of supererogation I will discuss three authors and their versions of the problem of supererogation: Dancy (1993), Nagel (1986) and Raz (1975a,b). Dancy's theory is an example of how the problem of supererogation may be solved when the agent acting supererogatorily respects the balance of reasons. Nagel is an example of solving a version of the problem of supererogation by going against the balance of reasons. Raz has a more peculiar position, because his solution to his version of the problem of supererogation involves respecting a relationship established between reasons (of different types), but this relationship cannot be properly described as a 'balance' of reasons.

3.2. The Problem of Supererogation and its General Form in Dancy's *Moral Reasons*

Supererogation is rarely discussed as a topic in its own right.⁶¹ More commonly supererogation is discussed because it presents a difficulty for a certain moral theory. I believe that the most interesting treatments of supererogation and its problems come embedded in a larger attempt at establishing a larger point about moral action. One consequence of this state of affairs is that the problem of supererogation (how can one explain very good deeds being deemed morally optional) will be transformed into various different problems due to the various purposes of each theoretical enterprise. The problem of supererogation, as common sense presents it, will be transformed into another difficulty, usually the difficulty that compelled the author to take supererogation into consideration. Another consequence of this state of affairs is that “the” problem of supererogation may be transformed into several problems within one and the same theoretical approach depending upon how general one's point of view is: it can reflect the larger purpose of the author or a narrower preoccupation. For example, it can ask a question about the relation between the deontic and evaluative domains, or a question about how to integrate the cost of an action for the agent into the overall balance of reasons.

In Dancy's *Moral Reasons*, the general declared aim is to find a “theory of motivation that will re-establish the possibility that moral reasons are both cognitive and able to motivate in their own right.”⁶² In elaborating a cognitivist and particularist

⁶¹ Notable exceptions: Heyd (1982) and Mellema (1991).

⁶² Dancy (1993, p.x)

theory of reasons for action, Dancy uses the metaphor of the “shape” of a situation (i.e. what we perceive in a situation as relevant for action), and this is where supererogation and its problems becomes relevant. Dancy maintains that in assessing any situation as agents we take into consideration certain features of the situation, features that appear to us as relevant or important for our actions. He calls the configuration formed by these features *salience* or *shape*. This is because the relevant features of a situation do not form a mere list: their order and their interconnections matter. They can enhance or weaken one another's importance, or they can cancel each other or come into conflict and remain in tension, for example.

Supererogation is discussed at this point because it poses a problem for the metaphor of shape. One situation is supposed to present us with only one kind of shape. The shape is identified by Dancy with the thin properties presented by that situation: the salient features are supposed to point to what is right or wrong, good or bad.

When I first considered the notion of a thin property, I ended by identifying a thin property with the shape in which the thicker properties of the case presented themselves. But now we see that there are two different sorts of thin properties, the evaluative and the deontic, that the same situation can generate one of each sort (by both calling for a certain action and permitting one not to do it), and that the two can vary apparently independently. This places pressure on the notion of shape, because I seem now to be committed to holding that one situation can have more than one sort of shape at the same time. (Dancy 1993, p.128)

The question of how / if the good and the right (i.e. the evaluative and the deontic) connect or diverge is a question of considerable proportions. Rawls, for example,

considers it to be *the problem* of any ethical theory.⁶³ Supererogation is relevant for only one aspect of this complex problem. For, if one accepts that there are such things as supererogatory actions (*de facto* or merely *de jure*),⁶⁴ then one also accepts that some morally good actions are not obligatory. A symmetry between domains might seem, after all, an intuitively attractive idea: that what is bad is forbidden, what is indifferent is permitted and what is good is obligatory. Even if one does not adopt sharp distinctions and prefers to see things as a matter of degree, it still seems intuitive that the worse an action is, the more likely it is to be under an interdiction and, symmetrically, the better an action is, the more likely it is to be deemed obligatory. Supererogation calls into doubt the symmetrical assumption: that the better something is, the more likely it is to be deemed morally obligatory. Moreover, it is the point where we seem to accept that some morally excellent deeds are not and cannot be obligatory.

In describing the commonsensical view of supererogation, Dreier (2004) and Hurka and Schubert (2012) distinguish two sides: the optional side (the permission to act or not in a supererogatory manner) and the “going beyond duty” side (that supererogatory deeds are better moral deeds). If these two sides are seen as each representing one domain, namely the deontic domain and the evaluative domain, then the commonsensical problem of supererogation (how can one be permitted not to do

⁶³ In *A Theory of Justice*: “The two main concepts of ethics are those of the right and the good ; the concept of a morally worthy person is, I believe, derived from them. The structure of an ethical theory is, then, largely determined by how it defines and connects these two basic notions. Now it seems that the simplest way of relating them is taken by teleological theories: the good is defined independently from the right, and then the right is defined as that which maximizes the good.” (Rawls 1971, p.24)

⁶⁴ What I mean by *de jure* is that position according to which the idea of supererogation is coherent, but there are no actual supererogatory deeds.

some morally excellent deeds) is transformed into a problem of connecting the two domains: how can an increase in goodness not be followed by an obligation to act? And if the two domains do not increase in step with one another, then what exactly is the relation between them? This is the more general context in which Dancy raises the problem of supererogation:

Such acts characteristically enjoy a very high degree of value, probably more value than any other act available to the agent. But in thinking of them as supererogatory, we are thinking of them as actions which is not wrong of the agent not to do. The puzzle about supererogation is how there can be such acts. (Dancy 1993, p.127)

Dancy is clear about the fact that the commonsensical view of the problem is changed so as to accommodate the specific theoretical context, by choosing a notion of 'ought' that is more general than the one of 'duty' and by framing the problem as a problem regarding the relation between the evaluative and deontic domains:

As I have expressed it so far, the concept of supererogation involves a contrast between duty and value, suggesting to us that duty can give out where value persists and grows. But in fact this is too narrow a perspective on the problem. We should not focus too directly on the notion of duty, since there may be actions which one *ought* to do but which can not exactly be said to be one's *duty*. In saying that the agent is not wrong not to do such action, we are leaving behind the notion of duty and working with a more general contrast between an action which ought to be done and one which has value. This is a contrast between the deontic and the evaluative properties of an action. (Dancy 1993, p.127)

At this general stage, Dancy has at least three formulations of the problem. They look quite similar and therefore I think they can be considered to be different formulations of the same general problem:

- a) In terms of duty compared to value: How can one explain that “duty can give out where value persists and grows”?⁶⁵
- b) In terms of evaluative properties compared to deontic properties of a situation: How can one explain that “as value rises, the change in degree of evaluative property” is not “attended by a change in degree of deontic property”?⁶²
- c) In terms of “ought” following the good: How can one explain that “the more good one's action would do”, it is not the case that “the more one ought to choose it in preference to others”?⁶²

Regarding the latter formulation, notice that supporting the idea that “ought” always follows the good is the usual way of arguing for an anti-supererogationist position. An anti-supererogationist will usually be committed to the so called “good-ought tie-up”,⁶⁶ namely the claim that when something is morally good or excellent, it should be placed by a moral theory under an obligation, under an “ought”. Therefore a classical way of affirming the existence or possibility of supererogation is to deny that there is such a requirement, of a “good-ought tie-up”.

These three formulations, even though different, having different degrees of generality, for example, are, nevertheless, quite similar. I believe that a more radical transformation of the problem takes place when making steps towards an answer.

⁶⁵ Dancy (1993, p.127).

⁶⁶ Heyd (2002/2016).

3.3. Agent-neutral and Agent-relative Reasons

In search for a solution to the problem of supererogation Dancy appeals to reasons.⁶⁷ His entire book (1993) is dedicated to the development of a theory of reasons for action, so there is no surprise that the solution to the conundrum will be thought in terms of reasons.

There are two main characteristics usually attributed to reasons, that are important here. First, they have “deontic force”, which is a metaphorical way of saying that they are thought to direct action in a rather imperative manner (as if carrying an “ought” of varying intensity); in this sense, when one has (decisive or overall) reasons for a certain action, then this is what what one should do. To quote Dancy (2004, p.91), “some reasons are obviously in the business of more or less telling us what to do”. In “Enticing Reasons”, Dancy calls this kind of reasons “peremptory” in their style of favoring action.⁶⁸ Illuminating considerations in this respect can be found in Little (2013, p.112):

On a certain classical picture, reasons, by their very nature, are deontic. If one faces a reason in favour of an action, then one would be *wrong* not to follow its lead, absent sufficient counter-veiling justification. Depending on the type of reason, the wrong in question might be the wrong of immorality, or imprudence, or just the generic wrong of practical reason (what I will refer to in this chapter as 'irrationality' or 'impermissibility'). (...) Still, to understand something as a practical reason necessarily brings with it a kind of deontic

⁶⁷ Each time I speak about reasons, I usually refer to moral reasons because the general topic of the discussion here regards almost exclusively moral reasons. This way of speaking should be distinguished from another claim, namely that Jonathan Dancy makes a sharp distinction between moral reasons and other kinds of reasons. His claim in *Moral Reasons* (Chapter 3) is that moral and non-moral reasons are different not in content but in style.

⁶⁸ Dancy (2004) argues that there are reasons that have a different style of favoring action, namely the “enticing” reasons.

vulnerability: one who faces such a reason now stands in need of adequate justification to do other than it directs, on pain of going wrong.

Indeed, for many the idea is virtually definitional of a practical reason. It is found, *inter alia*, in the view that it is wrong not to act on the 'balance of reasons', though it is not limited to a particular view of weightings. (...) Reasons for actions are normative entities inherently on their way toward being all things considered deontic oughts. Call this the deontic view of reasons.

To this first characteristic of reasons, of being “by their very nature” or by definition, deontic, a second one can be added that is also important.

The second important characteristic is that reasons are generated by value. Important for Dancy in his discussion of supererogation are *two kinds of values*⁶⁹ *generating two kinds of reasons: agent-neutral value generating agent-neutral reasons and agent-relative value generating agent-relative reasons.*

The contrast is not between the moral reasons for action, which all derive from considerations of value, and what we can get the weak-minded to do. It is between one sort of reason deriving from what we might call objective or agent-neutral value, and another which focuses on the agent-relative, but which still concerns value, and *moral* value, for all that. (Dancy 1993, p.139)

The distinction between agent-neutral and agent-relative reasons is relevant in discussing supererogation for reasons that I will explain below. The distinction itself is borrowed from Nagel (1986), who distinguishes between three categories of reasons: one that is purely subjective and two that have different degrees of objectivity. The ones that are deemed to be objective acquire this status due to their capacity to pass an “objectification test”. They are sub-divided into agent-relative and

⁶⁹ In Chapter 12 of his (1993), Dancy elaborates this claim and maintains that he speaks rather about different kinds of *valuing* than different kinds of *values*.

agent-neutral reasons. How one establishes objectivity is a rather complicated problem and Dancy discusses various conditions for objectivity and what exactly that might mean in a dedicated chapter in his book. I think it is enough, for present purposes, to say, roughly, that a reason has some degree of objectivity if it could be recognized as a valid reason by someone else than the agent. One can distinguish intuitively between reasons that are purely subjective (i.e. reasons that are reasons only for the agent and are not expected to have any validity or to be recognizable as reasons for anyone else – like the reason why I would choose chocolate ice cream over vanilla) and reasons that have some degree of objectivity. The highest degree of objectivity is present in agent-neutral reasons⁷⁰ (recognizable as such from perspectives very different from the agent's); agent-relative reasons, such as “because he is my friend” or “because she is my sister” are not reasons with the same standing for any agent (therefore they have less objectivity), but any agent can recognize them as valid reasons for someone standing in those relations to others.

Dancy's characterization of the distinction between agent-relative and agent-neutral reasons is the following:

Nagel holds that there are three sorts of reasons for action. The first are stubbornly subjective reasons, such as those which are in play when we are choosing from a menu in a restaurant. Here we have no desire that our reasons should somehow receive objective validation, and those reasons, depending as they do on our own peculiarities (our taste), cannot be recognized from a viewpoint at any great distance from our own. But there are two classes of objective reasons. The first are agent-relative reasons, of which we will be hearing a great deal; the second are agent-neutral reasons. Both of these are recognizable at some distance from here, as we shave peculiarities from our

⁷⁰ Nagel (1986), for example, takes the avoidance of pain to be an agent-neutral, highly objective reason, obvious to any agent as a reason. Of course, that does not mean it is considered a reason impossible to defeat (an “absolute” reason for Raz (1975a)).

perspective in the move towards objectivity. The agent-relative ones are less objective, of course, though they can be recognized, and in some sense endorsed, from more objective points of view. (Dancy 1993, p.146)

I have claimed that this distinction, between agent-neutral and agent-relative reasons is important when trying to explain supererogation. In order to understand its importance, let us suppose for the moment *that we can do without it*, that there is no distinction being made between kinds of reasons depending on their relation to the agent. What would that picture look like for supererogation?

Once moral reasons (with their deontic quality) have been introduced, there is no general threshold of obligation beyond which deeds become supererogatory. Instead, the agent is supposed to decide if something should be done or not in each particular case by taking into account a “balance of reasons”, i.e. by evaluating how considerations counting for and against doing the act relate and influence each other.

On the one hand, this makes the evaluation more flexible and capable of degrees: instead of establishing if something is a duty or not (simpliciter) one can establish various degrees for various situations of being morally more or less compelling. On the other hand, the problem of supererogation seems to have become more stringent in this new setting: if value generates reasons, then presumably high value or more value generates more reasons. And reasons are usually understood as deontic in nature, as carrying an “ought”. Therefore, wherever there is value, and value generates reasons, there is an “ought”. In this way, one is back at the old intuition that the ought should follow the good, that the highly valuable deed should be obligatory; in other words, one arrives at the good-ought tie-up.

The question is, therefore, how can supererogation be made possible in this setting? If each time something is morally good there is also an obligation to bring it about, then it seems that there is no place for a situation where one could rightly say that something is good but nevertheless no one could be obligated to bring it about (as we would plausibly say in the case of the hero giving his life to save others). The problem of supererogation has thereby morphed thereby into the problem “(...) *how to cope with the intuition that increased value generates increased reason to act*”.⁷¹

As I see it, there are two main strategies available. *First, one could deny that all reasons are “oughty”*, that they are all deontic by definition or by their “nature”. Maybe the reasons counting in favor of (what we usually consider to be) supererogatory actions are some other kind of reasons, a kind that does not have various degrees of deontic force implicit in their employment. The problem with the deontic nature of reasons is not that they always impose a full-scale obligation; rather, it is that they seem to impose some kind of justification or a countervailing force if one is not following them. If someone has a valid reason to do something, then simply not doing it is not necessarily wrong, but it does seem to leave the onlooker puzzled, in need of a certain justification why that reason was not followed even when it was recognized. Supererogation seems to allow exactly this: to not do a morally excellent deed without further need for excuse or justification. To repeat, a first possible strategy is to deny that the value of supererogatory acts generates *deontic* reasons.

A second possible strategy is to accept that all reasons involved in

⁷¹ Dancy (1993, p. 141).

supererogation are deontic, but to deny that they all pull in the same direction (i.e. they might be in opposition). Namely, it is to say that even though the good implied in the supererogatory deed generates a requirement, there are also reasons opposing this requirement and this is why, sometimes the agent is allowed to ignore the first requirement (i.e. to act supererogatorily) because on the balance of reasons, the opposing reasons sometimes might win. In other words, the strategy is to accept “the claim that if there is a reason in favor of doing something and no reason against, one ought to do it”⁷² and to go on to say that there is sometimes *reason against* doing something supererogatory, namely the reasons generated by the agent's sacrifice, *i.e. agent-relative reasons*. Dancy will consider the first strategy, but reject it in the end. He will use the second strategy when providing an answer to the problem of supererogation.

3.4. The Two Strategies for Responding to the Problem of Supererogation

In order to understand the first strategy, one should first ask :what kind of reasons could there be such that the agent would be permitted not to act upon them without further ado (because the reasons are non-deontic)? Two prominent examples are Kagan (1989), who considers briefly the idea of “non-insistent” reasons only to dismiss it, and Horgan and Timmons (2010) who propose and adopt “moral merit-conferring” reasons in order to solve their version of the problem of supererogation.

⁷² Dancy (2004, p. 110).

Kagan (1989, p.379)⁷³ characterizes non-insistent reasons as reasons which

(...) would have genuine weight, and might well be able to outweigh various other reasons. But they themselves would not be *insistent*, in that from the rational perspective it would be perfectly acceptable to neglect them. Presumably, by the very nature of the case, such non-insistent reasons could never ground a *requirement*. But, in contrast to the previous suggestion, this would be due to the intrinsic nature of the *reasons*, rather than merely verbal gerrymandering with the notion of ‘moral requirement’.

However he concludes that non-insistent reasons do not make sense and that all reasons are insistent even if they can be opposed and outweighed:

The insistent nature of reasons comes out even more clearly when they are considered in isolation. (...) There would be no rational justification for neglecting the sole reason: the reason is insistent. Yet why should the reason be any *less* insistent when it is opposed by countervailing reasons? It may be outweighed, to be sure, but it is difficult to see what could suddenly make it non-insistent. (Kagan 1989, p.381)

Horgan and Timmons (2010), on the other hand, want to deny that “If there are morally best reasons that favor performing an action that one is in a position to perform, then one is at least *prima facie* morally required to perform that action, and one is perhaps all-in required to perform that action.”⁷⁴ They do that by denying that reasons for actions may have only one kind of role to play, the requiring role. And they propose the "moral merit conferring" role for reasons:

(...) we propose what we call a nonrequiring, “moral merit-conferring” role that a moral reason can play and that makes sense of the very idea of supererogatory acts: acts that are completely morally optional, but

⁷³ Kagan (1989, p.379) analyses a possible case of non-insistent reasons (like reasons to pick one particular meal rather than another in a restaurant) but nevertheless rejects the possibility of such reasons: “I do not know what reasons of this kind would be like”.

⁷⁴ Horgan and Timmons (2010, p.36).

good and morally meritorious to perform. (Horgan and Timmons 2010, p.50)

The strategy, as I understand it, is to deny that "the only way in which a moral reason can favor an action is by tending to require that action."⁷⁵

Dancy (2004) himself proposes his own kind of non-deontic reasons, namely reasons that "entice" rather than require; the latter are "peremptory"⁷⁶ in style:

(...) I consider the possibility of a general distinction between two sorts of reasons: those that entice and those that are, as I will later say, more 'peremptory' in style. The main idea here is that some reasons are obviously in the business of more or less telling us what to do (...) Other reasons, however, are more to do with making an option attractive rather than demanded, required or right. At its most basic, the distinction I am after is between those reasons that say 'Come on, I'd be nice', and those that say something of a more forceful sort such as 'You'd better do this'. (Dancy 2004, p.91)

But even though the proposal withstands criticism, Dancy does not think that supererogation can be explained by the existence of enticing reasons because "The whole story of supererogation given above makes no essential appeal to non-peremptory reasons."⁷⁷ This is because none of the reasons involved in supererogation have usually to do with "the fun, the amusing, the attractive", like enticers typically do.

Therefore, Dancy's solution will be based on accepting the peremptory character of reasons involved in supererogatory action. His way way out from the

⁷⁵ Horgan and Timmons (2010, p. 62).

⁷⁶ In his (2004) article, Dancy also uses the term "non-insistent reason", but, I think, quite differently from Kagan. For Dancy peremptory reasons can be non-insistent in the sense that they can be discounted by the agent. By contrast, for Kagan, had there been such things as non-insistent reasons, they would be non-insistent by their very "nature".

⁷⁷ Dancy (2004, p.114).

good-ought tie-up is given by the distinction between agent-neutral and agent-relative reasons, i.e. what I have called a *second* possible strategy.

3. 5. Dancy's Solution to the Problem of Supererogation

This *second strategy* involves accepting that the reasons favoring supererogatory action are peremptory, but states that the agent might have *her own reasons*, namely agent-relative reasons with moral standing, that can count as reasons against doing the supererogatory deed. This strategy uses, a clause embedded in the conditional that establishes the “oughtiness” of reasons, namely in “the claim that if there is a reason in favor of doing something *and no reason against*, one ought to do it”. It is the “no reason against” clause that is at issue here. This strategy claims that there might be moral reasons (i.e. *agent-relative reasons* stemming from the sacrifice of the agent) going against the objective reasons favoring a certain supererogatory deed.

What this would mean, if we could get the story to run, is that increase in neutral value need not necessarily lead to increased reason to act or make it more wrong to fail to act. For increased neutral value can be counterbalanced by increased cost to the agent, i.e. by a decrease in agent-relative value. (Dancy 1993, p.138)

In other words, Dancy accepts that the good produced by a supererogatory action gives the agent reasons to act supererogatorily. However, he argues that the picture is much more complex, namely, more complex than the picture of an agent confronted with a difficult course of action required by moral reasons. Dancy argues

that the objectively good result of a moral deed, i.e. its objective value, is not the *only component* of the generated reason and its deontic force. There are two other components needed in order for a reason to be formed: a special relationship between the agent and the situation (e.g. opportunity – for some good deeds should be done but not necessarily by this agent, here and now) and the sacrifice of the agent. Therefore, the picture gains complexity:

(...) there is also a systematic relation between the deontic and the evaluative, which we expressed by saying that the deontic properties are resultant properties, and they result from a combination of:

- neutral value
- special relation
- sacrifice involved, generating agent-relative value (Dancy 1993, p.141)

Consider the example⁷⁸ of a doctor who can volunteer to go to another country to help fight against a dangerous epidemic. This kind of volunteering is obviously valuable; it is good for everyone else and it also involves some considerable risks for the agent. Measured objectively, from an impersonal point of view, the risk for one person (the doctor) seems to be a small price to pay for the good that might be done for many. The value generated for all by this gesture would seem to point to strong agent-neutral reasons in favor, which also means that - all things considered - we judge that the doctor *should* offer her help. Dancy stresses that the cost for the agent is already taken into account in this impersonal, value neutral picture. His proposal is to add to the neutral value picture *the same cost, but seen in a different manner* (assuming that the special relation is already in place). Namely, to see the agent's

⁷⁸ I am borrowing this example from Scheffler (1982).

sacrifice, not from a neutral, impersonal perspective, but from the personal perspective of the agent, as a risk *for her*. This agent-relative decrease in value generates agent-relative reasons favoring the opposing course of action, and by this, reasons capable of overturning the initial verdict, of a moral obligation to help. Should the risks for the doctor's life or health be considerable, then this new picture provides a theoretical explanation for why the doctor would be within her moral rights not to go. This does not mean that agent-relative reasons will always be able to justify *not* making a sacrifice; it only means that they are sometimes able to do this:

Now I want to say that the deontic results from the combination of three elements , value, special relationship and sacrifice involved for the agent. These three contribute in their different ways to the creation of reasons. The value that the act would create probably generates strong reasons in favour of doing it, for any person who stands in the requisite special relationship. But the cost that person will have to pay, which is already counted once during the calculation of value, counts in a different way when the prospective agent comes to consider what to do. So there is on one side the value derived from the difference it will make to the world that this action was done, which we can now call *neutral* value. But there is something else to be considered on the other side, and since it seems best to consider this something else as centered on something of value to the agent, we should think of this something else as value too, but as *agent-relative* value. (Dancy 1993, p.138)

Therefore, Dancy's solution to the problem “(...) How very great value can fail to create a requirement to act where there is opportunity”⁷⁹ is to allow agent-relative reasons to have a moral weight able to sometimes surpass (depending on the demandingness of the sacrifice) the weight of agent-neutral reasons. The “great value” is neutral value, the value impersonally considered; it does not automatically create a requirement because, in order to do so, one should take two other factors into account

⁷⁹ Dancy (1993, p.131).

(according to Dancy): a special relation to the agent (if the agent has an opportunity to act, for example) and the decrease in agent-relative value created by the agent's sacrifice.

This is the justification for our commonsensical intuition that one is not required to perform some morally excellent acts: because the sacrifice for the agent might count as a moral reason against the action.

So I want to say that the cost to the agent creates reasons for the agent that are distinct from its contribution (presumably negative) to the sum of neutral value. And these reasons are ones that we can all recognize and appreciate. This is what enables us to recognize that it is morally acceptable for the agent not to go by the balance of neutral value alone. We should not confuse this with some sort of special pleading on the part of the agent. (Dancy 1993, p.139)

There are two important comments to be made at this point. The first regards the moral value of agent-relative reasons. The second regards a further complication created by this answer.

It is important to note that Dancy attributes *moral* value to agent-relative reasons. This is a departure from the usual view. Usually, the demandingness of the sacrifice involved in acting supererogatorily is seen as a possible reason against doing the action, but only as a *prudential* reason against the action, i.e. without moral standing against other moral reasons. In other words, it is thought that the partiality and even selfishness of the agent is the one speaking against acting supererogatorily (while supererogatory action is understood as being for the common good). And this is why agent-relative reasons are usually dismissed as being “tainted” by partiality, and incapable of being veritable bearers of normativity. One might, for example,

recall Kagan's stern position in this respect, namely that even when the sacrifice demanded by morality is quite large,⁸⁰ the difficulty of the sacrifice does not represent a valid moral excuse (never mind a valid reason):

Morality requires that you perform—of those acts not otherwise forbidden—that act which can be reasonably expected to lead to the best consequences overall. Few of us believe this claim; none of us act in accordance with it. Consider just how radically demanding it is. It bids us to act not with an eye to merely furthering our own projects and interests, or those of some individuals we may favor—but with regard for the interests of all individuals, the world as a whole, overall good. It demands that I ask how I can make my greatest possible contribution, all things considered—even though this may impose considerable hardship on me—and it forbids me to do anything less. If the claim is correct, most of my actions are immoral, for almost nothing that I do makes optimal use of my time and resources; if I am honest with myself I will recognize that I constantly fail to do as much good as I am able. (Kagan 1989, p.1)

By contrast, Dancy considers that the “shrinking hero” is not guilty of moral failure, and that the agent's partial reasons may very well count as *moral* reasons:

The cost to the self is shown to be a moral reason by the fact that we, as outsiders, take it that the shrinking hero is within his moral rights to hold back. The contrast is not between the moral reasons for action, which all derive from considerations of value, and what we can get the weak-minded to do. It is between one sort of reason deriving from what we might call objective or agent-neutral value, and another which focuses on agent-relative, but which still concerns value, and *moral* value for all that. (Dancy 1993, p.139)

⁸⁰ Nevertheless, Kagan (1989, pp.7-8) notes that this position does not commit him to saying that extreme sacrifices are required: “The moderate might be tempted to object that an agent can go overboard on sacrifices—exhausting and impoverishing herself—ultimately destroying her ability to make continual contributions to the overall good (however this is ultimately understood). In this way the extremist's position might be thought to be self-defeating: were it only more moderate in its demands, agents would—in the long run—be able to do the world more good. Such an objection, however, would rest on a misunderstanding of the extremist, who does not demand sacrifices for their own sake—but only insofar as they are the cost of producing the greatest possible good. The extremist would be the first to urge that mindlessly driving oneself to exhaustion or recklessly dispensing one's goods can be counterproductive. Far from being required, such behavior is forbidden.”

My second comment regarding Dancy's solution is that his answer to the problem promptly creates another problem, one that he fully acknowledges. For let us suppose that the risks involved in a doctor's volunteering abroad are so great that the balance of reasons comes in favor of not going abroad and therefore not acting supererogatorily. Now the doctor is morally withing her rights not to go; however, in this case it would also seem that it is wrong *to go*, because it would be against the balance of reasons:

What about the other side of the coin, which is the agent's right to ignore the cost to self and make the sacrifice? The standard problem here is that if we succeed in justifying the agent's failure to do the heroic act, we can no longer approve his choosing to be a hero. (Dancy 1993, p.140)

Dancy's response to this problem is that the agent (and the agent alone) has a moral right to discount the reasons going against her sacrifice, because it is an obvious ethical truth that one can discount the cost because it is oneself who incurs it:

Others, of course, do not have the right to discount the cost to the agent, whichever way of valuing we are talking about. (...) If anyone doubts this account and asks why the agent has the right to discount (not to ignore) the cost to self, it is hard to know what one could appeal to in an answer which is basic enough to offer an advance on what we have already. The only thing that I can think of to say is that the agent has this right because it is he that is going to have to pay the cost. (...) But without begging the question in favour of consequentialist ways of looking at things, no theory should be taken to be in difficulties on this point unless it itself serves to invalidate the apparently obvious ethical truth that I am appealing to here. (Dancy 1993, p.215)

I will discuss in the next section how this position affects the problem of supererogation.

3.6. Conclusion: Dancy about Supererogation

Dancy's preoccupation with supererogation started with the discrepancy between two moral intuitions. On the one hand, it seems intuitive that the deontic and the evaluative domains should have a kind of symmetry, in the sense that the positive extreme of moral goodness should be associated with the positive extreme of moral obligation:

Normally we would think that, as value rises, the change in degree of evaluative property is attended by a change in degree of deontic property; the more good one's action would do, the more one ought to choose it in preference to others. (Dancy 1993, p.127)

On the other hand, it also seems intuitive that in cases of supererogation a very good action is and should be left optional, i.e. non-obligatory. At this first stage, the problem was formulated in very general terms, namely in terms of deontic *versus* evaluative: *how can one explain that "as value rises, the change in degree of evaluative property: is not "attended by a change in degree of deontic property."*⁸¹

The answer to this problem required some theoretical elaboration and therefore led to a second, less general view on the problem of supererogation. Once reasons are introduced, the image of the deontic domain becomes more complex and can account for various forces pulling in different directions. Supererogation was explained by appeal to two kinds of reasons, the agent-neutral (favoring the action due to the objective goodness to be obtained) and the agent-relative (favoring

⁸¹ Dancy (1993, p.127).

refraining from action due to the sacrifice on the part of the agent). I believe that, the progress towards Dancy's solution can be described step by step as follows:

- a) If a supererogatory action is good, then it has value.
- b) If the supererogatory action is good for all, impersonally, then it has neutral value.
- c) If it has neutral value, then there are agent-neutral reasons to perform the supererogatory action.
- d) All reasons are deontic entities; they carry normative force.
- e) Therefore, if there are agent-neutral reasons to do the supererogatory action and no reasons against it, then there is a kind of moral obligation to act supererogatorily.

In order to avoid this problematic conclusion, two strategies were explored: to reject d), namely that all reasons are “oughty” or to accept d) and to deny that there could be no *moral* reasons against doing something supererogatory. To repeat, Dancy rejects the first strategy and adopts the second.

These considerations, about agent neutral and agent-relative value allow the formulation of a slightly different and more specific problem of supererogation. If the good brought about by the supererogatory act is considered to be of neutral value, and that value generates agent-neutral reasons, then the problem of supererogation is *'how can one explain that the agent is allowed not to follow the agent-neutral reasons?'*

Dancy's response is that agent-relative reasons, stemming from the sacrifice incumbent on the agent, may provide the necessary justification for the “shrinking hero” because agent-relative reasons may have moral standing (and may be discounted by the agent in case she decides to be a hero after all).

In the case of supererogation, what is to justify the reluctant hero in hanging back is the cost to him of sacrificing himself for the greater good. (...) Neutrally calculated, the cost is the same whoever is to pay it, and so there is a moral calculation which is blind to the question who is to pay. (...) The individuality of the agent has not come into the matter at all yet; which person the agent is bears no direct moral relevance to the neutral calculation. But that the cost is to be his cost makes a great difference to him, and here the identity of the payer makes a difference as well as his nature. The cost is counted twice, though not in the same way both times; perhaps it would be better to say that it has double relevance. (Dancy 1993, p.213)

Therefore, one may distinguish at least two slightly different problems of supererogation, depending on how general the discussion is taken to be. The first problem is 'how to explain that the increased goodness of a deed is not matched by an increase in deontic value (i.e. by the imposition of an obligation)'. The second is formulated in terms of reasons: how can one explain that the agent may legitimately choose not to follow agent-neutral reasons.

3.7. Thomas Nagel and the Coexistence of Personal and Impersonal Standpoints

Even though various problems connected to supererogation seem to pop up in very different theoretical environments, the most fertile terrain for supererogation trouble remains, I believe, the discussion about personal and impersonal aspects of

morality. The reason for this state of affairs is not obvious: nothing in the very concept of supererogation seems to be directly connected with the personal/impersonal distinction in ethics. However, the morally good deed presupposed by supererogation, together with the permission to omit it, typically bring to light a discussion about *options* and about when/if moral requirements are involved. And the discussion of options for the agent, especially the option to forego a good deed, tends to reopen the issue of impersonal demands of morality *versus* the personal side of allowing the agent to opt out from some impersonal demands. In terms of reasons, the agent-neutral and the agent-relative reasons will represent the battlefield of impersonal versus personal ethical standpoints.

Nagel's main theme in *The View from Nowhere* seems to have this tension at its core, between personal and impersonal standpoints. Nagel takes his task to be a conciliatory one:

(...) to describe a kind of reconciliation between the objective standpoint and the inner perspective of agency which reduces the radical detachment produced by initial contemplation of ourselves as creatures in the world. (Nagel 1986, p.126)

But what is the objective standpoint and why is it important in ethics? Nagel believes it is a sign of our humanity to be able to "step back from our motive" (p.127) and take a larger perspective, beyond our narrow interest. The subjective perspective is given, we start from it, and by successive reflective steps we can acquire a more objective one, as it were by climbing step by step to arrive at the next 'meta' level:

To acquire a more objective understanding of some aspect of life or world, we step back from our initial view of it and form a new

conception which has that view and its relation to the world as its object. (Nagel 1986, p.4)

As one climbs the ladder of objectivity it should become clearer and clearer that the self is part of a “world of which itself is not the center”, that its previously thought centrality makes room for a less flattering position. The objectivity and subjectivity of a point of view are, in this way, matters of degree. The succession of steps towards objectivity⁸² is supposed to get rid of the peculiarities and the particularity of one's point of view, such that it gradually becomes possible to take someone else's view and, at the limit, no-one-in-particular's view – it is the view from nowhere:

Though I shall for convenience often speak of two standpoints, the subjective and the objective, and though the various places in which this opposition is found have much more in common, the distinction between more subjective and more objective views is really a matter of degree, and it covers a wide spectrum. A view or form of thought is more objective than another if it relies less on the specifics of the individual's makeup and position in the world, or on the character of the particular type of creature he is. (Nagel 1986, p.5)

When this advancement towards objectivity takes place with respect to action and values (as opposed to theory and truth), what it produces is a new set of values, not a new set of beliefs:

In theoretical reasoning objectivity is advanced when we form a new conception of reality that includes ourselves as components. This involves an alteration or at least an extension of our beliefs. In the sphere of values or practical reasoning, the problem is different. As in the theoretical case, we must take up a new, comprehensive viewpoint after stepping back and including our former perspective in what is to be understood. But here the new viewpoint will be not a new set of beliefs, but a new or extended set of values. (Nagel 1986, p.138)

⁸² Dancy (1993) detects an ambiguity in Nagel's description of the “objectification” process and distinguishes two meanings for Nagel's claims. Only one of the two possible interpretations will be deemed fit for the field of ethics and its objectivity.

Of course, it is not possible to have a *completely* objective position, one that pays tribute to no perspective, the perfect view from nowhere. The process of gradually becoming more detached from one's own perspective (and therefore more objective) is “not a completable task” (p.129). This detachment, of the agent, is supposed to bring about the realization that one is just an individual among others, with moral standing equal to that of others: “The standpoint from which one sees oneself as just an individual among others, viewing one's interests and concerns entirely from outside” (p.134).

All this manifests itself in the formation of impersonal values, and the modification of conduct and motivation in accordance with them. It imposes serious constraints. Values are judgments from a standpoint external to ourselves about how to be and how to live. Because they are accepted from an impersonal standpoint, they apply not only to the point of view of the particular person I happen to be, but generally. They tell me how I should live because they tell me how anyone should live. (Nagel 1986, p.135)

Climbing the ladder of objectivity creates impersonal values, and it shows how impersonal values are even possible. As Nagel (1986, p.138) puts it, “If we can make judgments about how we should live even after stepping outside of ourselves, they will provide the material for moral theory.” However, Nagel admits that there are several ways in which things can go wrong in the process:

- a) First, in taking a more and more objective standpoint, *values as such* might appear illusory: “Values can seem really to disappear when we step outside of our skins, so that it strikes us as a philosophical perception that they are illusory” (p.141).

- b) A second, related problem, is the problem of establishing the possibility of objective values, namely to “decide in what way, if at all, the idea of objectivity can be applied to practical questions, questions of what to do or want” (p.138). This seems for Nagel to be the essential task regarding ethics: ethics has to be “saved” from the threat of subjectivity. What is at stake here is the *possibility* of an objective system of ethics, based on objective values: “I want now to defend the objectivity of ethics by showing how that standpoint alters and constrains our motives” (p.138)
- c) A third problem may come from the other end of the process, namely from subjectivity. Instead of acquiring distance from one's own point of view and acknowledging other perspectives as equally important, the importance of one's own perspective can be aggrandized and exaggerated: “I acknowledge the dangers of false objectification, which elevates personal tastes and prejudices into cosmic values” (p.143).
- d) The forth kind of problem regards the clash between the objective and the subjective point of view. Nagel admits that, from a sufficiently large 'distance', personal issues that we care about will not look important anymore. In a way that was the whole point of the objectification process, to put things into perspective: some of our concerns will look smaller and some will disappear this way. However, not only the unimportant concerns may disappear; the important ones may be diminished, too:

It may be that from a standpoint sufficiently external to that of ordinary human life, not only chicken salad and salami but much of

what is important to human beings - their hopes, projects, ambitions, and very survival – cannot be seen positively to matter. (Nagel 1986, p.131)

Nagel indicates remedies for each of the four kinds of problems. However, I will only discuss the last kind of problem (the fourth) because this is where supererogation will enter into the picture.

Nagel's remedy for the fourth kind of problem lies in trying to keep a fine balance between the personal and the impersonal standpoint. Our deeply human concerns, which are also deeply personal (so they find themselves at odds with the impersonal perspective) are kept in the picture by what Nagel calls “objective toleration”. The main idea is that the objective point of view can recognize personal reasons as legitimate reasons to have, as long as they have limited claims:

(...) a strategy of objective tolerance as opposed to objective affirmation – is to find grounds for acting within my personal perspective that will not be *rejected* from a larger point of view: grounds which the objective self can tolerate because of their limited pretensions to objectivity. (Nagel 1986, p.130)

The strategy of objective tolerance is, as I see it, Nagel's way of making peace (or of running together the two kinds of values in a manner that does not require choosing between the two) between the two sides, namely the personal and the impersonal standpoint.

Some grounds for actions, claims Nagel, will not be *positively* endorsed when seen in an objective light, because they are not and cannot be *anyone else's* ground for action. However, they can be tolerated from the objective point of view because it is true, *objectively* true, *that* they are subjective grounds for action. In the same way as a

truth can be told about a lie, something objectively impartial can be said about partiality – this seems to be Nagel's point. But this is not the whole point because this can be said about grounds for action that cannot withstand objective scrutiny and therefore cannot be objectively tolerated. An inappropriate impulse of anger, for example, can be identified on reflection, as the subjective ground for a former action, but it will not be a “subjectively legitimate ground for action” and it cannot be the object of toleration. Toleration of some subjective grounds for action is needed because otherwise (if those grounds would be eliminated by objective detachment) some actions would become incomprehensible. Some actions (and values) can be understood *only* when inhabiting a subjective perspective, Nagel claims, and this why the objective point of view cannot eliminate them, on pain of making important parts of our human agency incomprehensible.⁸³

From the objective point of view my preference for chocolate over vanilla does not make much sense, it cannot be seen as a general ground for action (not even for my own action in general; it only makes sense for-me-when-choosing-ice-cream, i.e. in very particular circumstances). But it is rational to tolerate it (and therefore include it among my subjective reasons) from a detached point of view, if I do not claim it to have a higher significance than one particular quirk of taste. The problem becomes more complicated when the object of objective toleration is not taste in food, but “hopes, projects, ambitions”. They pass the test of objectification, according to Nagel, because the actions justified by hopes, projects, ambitions “will be fully

⁸³ This reminds one of the saying that you have to walk in someone's shoes to really understand them.

comprehensible only from a particular perspective within the world” (p.140), and therefore they will be deemed “subjectively legitimate grounds for action” (p.131).

One consequence of the objective toleration strategy is that agent-relative values and agent-relative reasons are seen as legitimate grounds for *moral* action. Nagel makes room for the personal side of ethics, namely for agent-relative values and agent-relative reasons:

A second respect in which reasons vary is in their relativity to the agent, the person for whom they are reasons. The distinction between reasons that are relative to the agent and reasons that are not is an extremely important one. If a reason can be given a general form which does not include an essential reference to the person who has it, it is an agent-neutral reason. For example, if it is a reason for anyone to do or want something that it would reduce the amount of wretchedness in the world, then that is a neutral reason. If on the other hand the general form of a reason does include an essential reference to the person who has it, it is an agent-relative reason. (Nagel 1986, pp.152-3)

He believes that what people have *reasons to do* is based not only on impersonal values but also on agent-relative values: “ethics is not based solely on impersonal values (...) We can no more assume that all values are impersonal than all reality is physical” (p.163).

A consequence of Nagel's attempt to have a peaceful compromise is his argument that the three most common types of agent-relative reasons can be seen as *objective reasons* (they are less objective than the impersonal ones but more objective than the purely subjective ones) even though they are, indeed, relative to the agent. Their status as objective reasons guarantees that they are legitimate reasons to be taken into account alongside with the agent-neutral reasons:

- a) reasons of *autonomy* – “stem from desires, projects, commitments, and personal ties of the individual agent, all of which give him reasons to act in the pursuit of ends that are his own” (p.165);
- b) reasons of *deontology* (somehow misleadingly named so) – “stem from claims of other persons not to be maltreated” (p.165);
- c) reasons of *obligation* – that are actually reasons of special obligations one has towards parents, children, spouses.

Another aspect of his peaceful attitude and compromise between sides is Nagel's view that from the impersonal point of view itself, the personal point of view has some justification, its own rational place. As a consequence the impersonal claims are not limited simply by the force of the personal claims, but they suffer a kind of limitation from within:

But I believe an answer can be sought from the impartial standpoint of morality, which will give to everyone a dispensation for a certain degree of partiality – in recognition of the fact that it is only one aspect of the human perspective. Like reason, the moral standpoint should try to recognize and explain its own limits. (Nagel 1986, p.202)

All these elements – the objective toleration strategy, recognizing the objectivity of agent-relative reasons, the objective standpoint limiting itself and giving dispensations for degrees of partiality – are elements meant to ensure a peaceful cohabitation between parts. This picture, of a compromise between impartial and partial standpoints, is one which is meant to be close to actual ethical decisions made by agents.

However, Nagel is quite aware that the peace cannot last, and that the conflict can be only partially prevented by these moves.⁸⁴ The conflict becomes particularly acute in his discussion of how the impartial/impersonal standpoint may drastically affect the achievement of a good life for the agent. Nagel claims that “It is clear that a strongly impersonal morality, with any significant requirements of impartiality, can pose a serious threat to the kind of personal life that many of us take to be desirable” (p.190). And this is because impersonal morality may be at times very demanding, claiming resources that the agent might otherwise prefer to dedicate to his own plans. Nagel believes this is an every day occurrence in our moral lives: any purchase of clothing, wine, theater tickets, vacations, gifts, books (his examples) may be said to be at fault from the point of view of an impersonal morality because the money used to buy them could bring, for example, famine relief in other parts of the world. And this is the domain of beneficence, where supererogation enters the discussion naturally.

3.8. Nagel on Supererogation

When considering the demands of impersonal morality against the personal sphere of the agent one is placed before a dialectical arena of grand proportions. The main question is how the agent's resources should be allocated: exclusively in

⁸⁴ Nagel (1986, p.185) writes: “This is a true philosophical dilemma; it arises out of our nature, which includes different points of view on the world. (...) I believe the human duality of perspective is too deep for us reasonably to hope to overcome it. A fully agent-neutral morality is not a plausible human goal.”

accordance with the public, impersonal good, meant to benefit everyone, *or* according to a more nuanced way of seeing things, that would allow the agent to sometimes go against the impersonal way of seeing the good. Two main ways of going against the impersonal good are usually taken into account: what has been called “options” and “constraints”. Options are similar to Nagel's reasons of autonomy and reasons of special obligation- they are permissions granted to the agent to not act in accordance with the general good, when important personal plans or other important human relations would be at disadvantage. Constraints are similar to Nagel's reasons of deontology – in that they forbid the agent to harm someone even if an impersonal good would be obtained this way. Positions like Slote's (1984) and Scheffler's (1982) want to keep the image of morality as based on a calculation of impersonal good, but they also want to design a system closer to commonsense morality, that would not accept extreme sacrifices for the common good to be a matter of moral obligation. Therefore, they introduce various kinds of options for the agent, such that, for example extreme sacrifices would not be demanded anymore. Voices like Williams (1981), Wolf (1979) or Railton (1984), while they differ among themselves, would rally against the idea that the demands of impersonal morality have an *a priori* prevalence when compared with personal standpoints. They consider the demands of impersonal morality as alienating for the moral agent.

In this grand battle over how far the demands of impersonal morality can go, supererogation is a small but central point. Those demands of impersonal morality that are deemed to be exaggerated, too demanding, are usually what common sense morality describes as “supererogatory” (because they involve an action with a morally

excellent result involving a serious sacrifice on the part of the agent). In a way, the whole dispute is about the possibility of supererogation.

My main task in this section is to establish where Nagel is situated in this ample dispute. This is not an easy thing to do because Nagel is trying to walk a fine line between personal and impersonal standpoints in order to leave the two in what he considers to be their natural state: in tension. His path to this result is quite sinuous, but I will try to describe it step by step. First, he would seem to have a standard moderate position by accepting both standpoints, personal and impersonal, as legitimate moral consideration. But he cannot maintain this position because his whole enterprise, from the start, is a search for the impersonal viewpoint, a search for “the view from nowhere” as a guarantee of unbiased truth and objectivity. Therefore, in the end, he returns to his preference to the objective standpoint and this where things start to look inhospitable for supererogation. I will detail these steps in what follows.

To repeat, the first step is to accept the credentials of the personal standpoint. Nagel seems to be one of the authors who would agree that the agent is justified sometimes in going against the recommendations of the impersonal good. He explicitly rejects the stricter line of argument, the one saying that the difficulty of the task imposed by impersonal morality does not justify reducing those requirements:

One might take the severe line that moral requirements result from a correct assessment of the weight of good and evil, impersonally revealed, that it is our job to bring our motives into the line with this, and that if we cannot do it because of personal weakness, this shows not that the requirements are excessive but that we are bad (...).

I do not believe this because even though morality has to emerge from

an impersonal standpoint, that standpoint must take into account the kind of complex beings for whom it is being devised. The impersonal is only one aspect of their nature, not the whole of it. What is reasonable to ask of them, and what is impersonally expected of them, should reflect this. (Nagel 1986, p.202)

From Kagan's (1989) critical point of view, for example, Nagel counts as a Moderate, namely someone who admits in his theory both options and constraints as exceptions from the requirements of the impersonal good and who would be, in this way, closer to commonsense morality. The usual reason given by a moderate is that the sacrifice required from the agent amounts to something far too demanding to be reasonable:

Many acts which would lead to the best results overall nonetheless are not required of us by ordinary morality, typically because the sacrifice would be too great to demand it of us. I am not required to devote my free time to fighting political oppression, nor must I give up my luxuries to support cancer research. On this view morality permits me to favor those things I most care about—whether my own welfare, the welfare of others, or any projects to which I am committed. (Kagan 1989, p.3)

From Nagel's (1986, p.202) claim that "it is unreasonable to expect people in general to sacrifice themselves and those to whom they have close personal ties to the general good", one might get the impression that indeed, according to Nagel, agent-relative reasons stemming from the personal side would have to sometimes defeat the agent-neutral reasons that an agent might have to follow the impersonal good. However, Nagel's position is more complicated than that. He wants to be able to claim some kind of universal superiority of the impersonal point of view, so *he is not willing to let the impersonal standpoint be defeated by reasons stemming from*

concerns about the well-being of the agent, even if he admits that the demands of the impersonal morality may drastically reduce well-being:

The moral life overrides the good life. (...) The idea is not that morality will necessarily conflict with the good life but that it can, and when it does it provides us with sufficient reason to sacrifice our own good. (Nagel 1986, p.196)

Nagel does not deny that sometimes the moral life (i.e. the impersonal standpoint) may be quite demanding, but he insists that even if the impartial morality is not the best choice regarding the agent's well-being, it is, nevertheless, *the best choice within the rational domain*.

My position resembles that of Wolf. She discusses the relation between the moral point of view and the point of view of individual perfection, and calls into question “the assumption that it is always better to be morally better”. (...) Where I part from her, as I explain below, is in my hope that there is some way of preserving the priority of moral requirements – if not of moral considerations generally – in determining how is it rational to live, though not how it is good to live. (Nagel 1986, p.197, fn.6)

Nagel does not claim that the rational superiority of impersonal morality is indubitable. It is for him something closer to a hope or an aspiration: “I am inclined strongly to hope, and less strongly to believe, that the correct morality will always have the preponderance of reasons on its side, even though it needn't coincide with the good life” (Nagel 1986, p.199). This aspiration makes sense because the whole discussion about ethics started with the intention of defending the *possibility of an objective ethics*. Objectivity, therefore, has to be the more valuable side of the two.

However, Nagel clearly feels the pull of the other side, fearing that, in these theoretical circumstances, some moral decisions would be rightly labeled irrational

for the agent to make. If the moral life always overrides the good life, this means that it should be so even in very extreme situations, when choosing to be moral comes at very high personal price. Nevertheless, choosing the impersonal moral choice *has to be rational*, indeed the best *rational* choice according to Nagel:

I have claimed that requirements of morality can conflict with the good life, and that this is not a legitimate ground for rejecting a morality. On the other hand, my own view is that the correctness of a morality is put in doubt if it would sometimes be irrational to accede to its demands (though the idea of such morality is not incoherent) . There is pressure from the moral standpoint itself to adjust those demands so that they converge with the condition of full human rationality, though not the condition of a good life. (Nagel 1986, pp.204-5)

Therefore, what Nagel wants to be able to say is that:

- A) The requirements of impersonal morality may conflict with the good life by demanding considerable sacrifices from the agent.⁸⁵
- B) It is always rational to choose the action indicated by impersonal morality over the indications of the good life.

The next section will detail the consequences of these tenets for supererogation.

⁸⁵ For example, Nagel (1986, p.197) claims: “While doing the right thing is part of living well it is not the whole of it, nor even the dominant part: because the impersonal standpoint that acknowledges the claims of morality is only one aspect of a normal individual among others. And there are times when doing the right thing may cost more in terms of other aspects of the good life than it contributes to the good life in its own right.”

3.9. Nagel, Rationality and the Limitations of Human Nature

Once he has established that the impersonal viewpoint has to win in matters of morality (to repeat, after walking a fine line between the two, Nagel ultimately takes sides in this way), Nagel's main worry is to tackle the problem of the *demandingness of the moral impersonal viewpoint*. In other words, it is going to be very costly for the well-being of the agent to be morally required to obey the impersonal demands of morality. In some cases, Nagel, worries, it is going to be so costly for the agent, that the moral action may be deemed *irrational*. And this, he cannot allow. His way out is to make the *very costly moral actions* (i.e. what we usually call 'supererogatory') *optional*, in a rather complicated manner, namely by a “compromise between our lower and higher selves”.

In order to have both A) and B) , i.e. in order to say both that morality is demanding *and* that it is nevertheless always rational to choose the morally better option, Nagel has to be able to claim that choosing the impersonal morality over the good life is unlikely to be seen as *irrational*.⁸⁶ In other words, Nagel has to find a way to prevent the situation where the sacrifice demanded by impersonal morality will be so drastic that it cannot be rational for the agent to have *an obligation* to obey the demand. The strategy that offers a way out of this dilemma, thinks Nagel, turns some requirements into mere options, this being a “modification to accommodate the normal limitations of human nature”.

It is an attempt to make the best of an unsatisfactory situation. Insofar

⁸⁶ It is not clear to me if Nagel takes into account a possible difference between *unreasonable* and *irrational*. He does not mention such a distinction explicitly in *The View from Nowhere*.

as it reduces the requirements of impersonal morality, this will reflect an attitude of tolerance and realism about human nature, rather than the conviction that to act on a more demanding requirement would be irrational or wrong. The idea is that certain demands on the ordinary individual – to overcome his own needs, commitments, and attachments in favor of impersonal claims that he can also recognize – are unreasonable in a way that can be impersonally acknowledged. (Nagel 1986, p.203)

The reduction, therefore, consists in weakening some impersonal requirements to make the action optional. However, this does not mean that making the demanded sacrifice (i.e. following the impersonal recommendation) would be irrational. The fault of 'irrationality' seems to reside, for Nagel, not in the task *per se*, but in its deontic status, in *being required*. Of course, when an extraordinary good deed involving sacrifice is merely optional, not required, then this is a clear cut instance of supererogation. Supererogation is not only a fitting example for Nagel, but also the proof that common sense already has such provisions and modifications in place, because they are the natural way to proceed:

But this does not mean that it would be irrational for someone who can do so to accept demands, or rather to impose them on himself. So this aspect of impersonal morality may throw light on the puzzling subject of supererogation. Supererogatory virtue is shown by acts of exceptional sacrifice for the benefits of others. Such acts are praiseworthy and not regarded as irrational, but they are not thought to be either morally or rationally required. What is it about them that makes them good, indeed exceptionally good, and provides a reason for doing them, without at the same time providing a reason against not doing them that would make such a failure rationally unjustified and bad? (Nagel 1986, p.203)

Nagel's formulation in the above quote looks quite similar to the commonsensical formulation of the problem of supererogation. But, I will argue, it is

a deceptive similarity. Here Nagel seems to put the usual worry about supererogation in terms of reasons: “*What is it about them that makes them good, indeed exceptionally good, and provides a reason for doing them, without at the same time providing a reason against not doing them that would make such a failure rationally unjustified and bad?*” One might notice that there are *two* related questions being asked here:

- a) what makes supererogatory actions good and provides reason for them; and
- b) what makes supererogatory actions the kind of good actions that can be omitted, without the omission being rationally unjustified or bad.

My point is this: the generic form of the problem of supererogation (i.e. its commonsensical, imprecise formulation) worries about the second question, namely about b) – because it worries about the omission of a good act not being imputable. The question of how one can explain that an excellent moral deed is not obligatory, is actually, a question about what consequences are appropriate (if any) in case the agent chooses to omit the action. That is because, usually, obligation is understood as something that cannot be ignored without consequences.

By contrast, Nagel's main worry is the first question, namely point a) 'what makes supererogatory actions good and provides reason for them'; only then, as secondary, is point b) taken into discussion. I believe this is so because Nagel's main worry at this point is the *irrationality* of some highly moral good deeds (that are being required by the impersonal point of view). He claimed that impersonal morality can have a high cost and, at the same time, that it is always rational to choose

morality. He is concerned with defending the latter claim here. His point is to show that common sense does not consider supererogatory acts irrational. According to common sense, acting supererogatorily is optional, but one may go on and be a hero and that will be *laudable, not irrational*. Nagel believes this illustrates his thesis that doing what the impersonal standpoint indicates as good, is usually considered rational (and laudable),⁸⁷ even in extreme cases. The paragraph mentioning supererogation as an example begins by saying “But this does not mean that it would be irrational for someone who can do so to accept demands, or rather to impose them on himself.”⁸⁸

Therefore, I think that Nagel's version of the problem of supererogation is, actually, 'how can one explain that following impersonal morality by doing something with high moral value that involves serious sacrifice (and thereby hurting the well-being of the agent) is most of the time *considered to be rational* – as one can see in the common employment of the concept of supererogation?'

3.9.1. Nagel's Answer to the Problem of Supererogation

His answer to this problem is a bit convoluted. I think Nagel takes the commonsensical status of supererogation as a premise. He seems to claim that *because* one usually takes supererogatory action to be rational, this means that the action was not irrational or in other words, not done in spite of reasons to the contrary. This means that, according to Nagel, heroic (for example) action is *not done*

⁸⁷ See Hampton (1993) about sacrifices that are not rational or moral.

⁸⁸ Nagel (1986, p.203).

by contradicting a balance of reasons inclining towards the personal side (on the side of agent's well-being). In other words, if we take heroic action to be reasonable it is because it conforms to reasons, namely agent-neutral reasons given by the impartial morality. Let me explain this in more detail.

According to Nagel, there are three possible situations in a confrontation of opposing reasons: the *reasons against* the action are decisive (and therefore the action is irrational) or the *reasons for* the action are decisive (and therefore the action is rationally required) or neither kinds of reasons are decisive and, in this case, the action is neither irrational or required:

Where a number of different and opposing reasons bear on a decision, there are three possible outcomes with regard to the rationality of the alternatives. Either the reasons against may be decisive enough so that the act would be irrational; or the reasons for may be decisive enough so that the act is rationally required; or there may be enough reasons both for and against so that although the act is not rationally required, it would not be irrational either (...). (Nagel 1986, p.200)

One might think, following his characterization of supererogation, that supererogation will be placed by Nagel in the third, non-decisive case, because acting supererogatorily is neither irrational nor required. However, according to Nagel, this is (surprisingly) incorrect. He argues that this is not the case because the two sides (personal and impersonal) do not have, in the case of supererogation, relatively equal weight. And they do not have equal weight because one obviously has reason on its side: "After all, someone who does not make these sacrifices is failing to do something that he does have a morally estimable reason to do. Morality is not

indifferent between his doing it and his not doing it”.⁸⁹ Supererogation cannot be a case in which the balance of reasons inclines towards the personal side, either. If it were so, then abstaining from the heroic act would be required and doing the heroic act would be irrational. But we usually do not take heroic acts to be irrational or, for that matter, believe the omission of the act to be required. Therefore, the absence of the requirement to behave heroically cannot be attributed to the personal side winning the battle of reasons. That is to say, if there is no requirement to act supererogatorily, this is not because the reasons for acting supererogatorily were defeated by opposing reasons.

There remains just one other possibility (out of three listed above): that supererogation is a case of reasons inclining towards the heroic action. How can one explain, then, that the heroic action is not required? Nagel considers that this is *not a matter of reasons defeating other reasons, but a matter of objective tolerance*, a tolerance that makes us concede that extraordinary moral acts involving sacrifice cannot be required from agents, even if they are what impartial reasons require.⁹⁰ Therefore, the heroic action is rational in spite of being harmful to the good life of the agent because: first, it is in accordance with the balance of reasons; and second, the requirements presented by these reasons have been modified and reduced to accommodate “the normal limitations of human nature” thereby resulting in the optional status of the heroic action:

(...) supererogatory virtue is adherence to the claims of impersonal morality prior to their modification to accommodate the normal

⁸⁹ Nagel (1986, p.204).

⁹⁰ I'm afraid that I agree with Dancy's criticism (1993) that this position sounds almost as saying that there are “requirements that do not require”.

limitations of human nature. This modification takes the form of a relaxation of these requirements through tolerance, as it were, rather than the discovery of new moral reasons that outweigh the original impersonal ones. If they had been outweighed, then there would be reasons against the type of sacrifice that displays supererogatory virtue: it would be wrong. As things are, it is merely not required. And those who undertake it nonetheless are praiseworthy for submitting themselves to the true strength of reasons that they could not reasonably be required to follow strictly, given the mixed character of human motives. (Nagel 1986, p.204)

There are two comments that I think are important for understanding Nagel's attitude towards supererogation. First, that the problem of supererogation is, not surprisingly, *split into two successive problems*. The first problem (and most important for Nagel) is *how can one explain that doing a heroic action is not irrational* (heroic action, for example, may endanger the well-being or even the life of the agent; being required to perform such a sacrifice may be regarded, therefore as irrational). The second problem stems from the solution given to the first: if one accepts that heroic action is not irrational because it aligns with reasons given by the impersonal side, then *how can one explain that heroic action is not required*. Only this, latter problem closely resembles the commonsensical version of the problem of supererogation. The former problem is nowhere on the radar of the commonsensical view and this is why I think it can be considered a different (even though related) problem of supererogation.

My second comment is that I do not think Nagel's appeal to supererogation works as intended. He seems to regard supererogation as a vindication of his point by commonsense morality, namely that usually we do not regard doing the supererogatory action as irrational (even though it is difficult and not required). This

much is true. However, I think he might have used a different concept of supererogation (different from the commonsensical one, that is). It is true that, according to common sense, supererogatory action is not irrational, and it is not required. However, Nagel's way of presenting supererogation as "not required" is quite peculiar. For Nagel, heroic action is not required just because of the "weakness of human nature". It sounds as though one actually *should* behave supererogatory but is allowed not to, not as a legitimate permission, but as an excuse.⁹¹ In the commonsensical picture *one does not need to be excused* for not acting supererogatorily.⁹² Nagel does not speak consistently in this manner. Sometimes it sounds as if objective toleration is a legitimate limitation of the requirements imposed by the impersonal standpoint. Moreover, Nagel sometimes claims that the impersonal standpoint limits itself and recognizes the *legitimacy* of the personal side's claims. If the impersonal standpoint recognizes the claims of the personal standpoint as legitimate, then nothing in this picture implies (yet) the inferiority of the standpoint being recognized as legitimate (on the contrary, it would seem an acceptance of equality regarding importance or at least, legitimacy). However, Nagel employs, at times, expressions that leave no doubt about the fact that he considers the claims of the impersonal point of view as more desirable and superior to the other point of view. The personal point of view is being accepted into the picture as a compromise

⁹¹ A point was made to me in conversation at the Virginia Philosophical Association (Oct. 2015) that how Nagel sees our "human weakness" looks less like an excuse, and more like general amnesty

⁹² My interpretation of Nagel's point of view is, again, not far from Dancy's (1993). It seems to me that "normal limitations of human nature" have the role of an excuse in front of what would otherwise be legitimate demands of impersonal morality. Even if we "strike a bargain between our higher and lower selves in arriving at an acceptable morality", as Nagel wants, the "lower self" is the one concerning the personal side, the one we should be embarrassed about. It is a compromise, not the way things should be. But this is not what supererogation is about. Intuitively, in clear cases of supererogation no excuse is needed if someone decides not to act supererogatorily, because the omission is not wrong.

because it represents our “weaker” side of humanity, our “lower self”, in contrast with the purity of impartiality:

We must so to speak strike a bargain between our higher and lower selves in arriving at an acceptable morality. In this way the gap between moral and more comprehensively rational requirements can be narrowed. It means that there is an impersonal sanction for striking the balance between personal and impersonal reasons in a certain way. (Nagel 1986, p.202)

In other words, heroic action is not required because humans are weak, not because it should not be required. And this line of thought is made even clearer by Nagel's claim that moral progress consists in accepting more and more impersonal claims, by compromising less with the personal side. Of course, this way of ranking the personal and the impersonal is understandable because Nagel's project is being designed from the start as a defense of the *objectivity* of ethics. Nevertheless, I think his argument based on the commonsensical view of supererogation fails because *his is not the commonsensical view of supererogation*.

3.10. Conclusion: Nagel on Supererogation and Its Problems

The main idea in *The View From Nowhere* is the confrontation between the personal perspective and the neutral perspective. Thomas Nagel is, without doubt, aware of the standoff between what he calls "personal and impersonal claims" in moral activity. As I said before, he does not even accept that this thesis, of the confrontation between the two, is only a possible frame of thinking. For Nagel, this is

just the reality of the lived moral life, a problem that confronts us no matter what we believe. His solution to this problem is a kind of compromise between “the objective and the subjective standpoint”, namely the method of “objective toleration” because “it is unreasonable to expect people in general to sacrifice themselves and those to whom they have close personal ties to the general good”.⁹³ For Nagel, the two sides do not necessarily exclude each other: one can see the justification for “personal claims” even if one takes the impersonal point of view. Therefore from the impersonal point of view itself, the personal point of view has some justification, its own rational place. As a consequence, for Nagel, the impersonal claims are not limited simply by the force of the personal claims, but they suffer a kind of limitation from within: “Like reason, the moral standpoint should try to recognize and explain its own limits”.⁹⁴ Supererogatory action is, for Nagel, a manifestation of the objective, impersonal standpoint whenever this standpoint is not limited by tolerance for human weakness.

The main point in invoking supererogation is to show that moral action (even the one implying heavy sacrifice) is not irrational. Therefore Nagel's version of the problem of supererogation is to ask *how one can justify that supererogatory action involving harm to the agent is not deemed irrational*. His answer to this problem immediately gives rise to another problem (hence my claim that the problem of supererogation is actually split into two different problems): *if heroic action is rational because it obeys impartial reasons, then how can one explain the fact that it*

⁹³ Nagel (1986, p.202).

⁹⁴ Nagel (1986, p.203).

is optional? The answer to this second question is to weaken requirements by objective toleration in order to accommodate “human weakness”.

One may notice that Nagel's solution to his problem of supererogation *falls outside the balance of reasons*: heroic action is not required *not* because there are reasons against it, defeating the reasons that would require it; but because objective toleration intervenes as a strategy from outside the balance of reasons and makes heroic action merely optional (to accommodate the demands of the personal side).

3.11. Raz and Supererogation

In *Practical Reasons and Norms*, Joseph Raz set himself an ambitious aim: “a logical theory of practical conflict”.⁹⁵ Supererogation is not obviously connected to this task. Nevertheless, once one is reminded that the problem of supererogation can be seen as a conflict between what the agent has reason to do and what the agent is permitted not to do, presenting supererogation as a particular case of conflict between reasons makes sense. Previous positions about the problem of supererogation could also be seen in terms of conflict: between agent-relative and agent-neutral reasons (Dancy 1993) or between personal and impersonal standpoints (Nagel 1986).

Raz has one of the clearest formulations of the problem of supererogation in terms of reasons:

One important characteristic of supererogatory acts (...) is that their performance is praiseworthy while their omission is not blameworthy. But this creates a problem. If doing a supererogatory act is

⁹⁵ Raz (1975b, p.36).

praiseworthy there must be reason for doing it, and the reasons must outweigh any conflicting reasons for not doing it. But if there are conclusive reasons for performing the act then not to perform it is to act against the balance of reasons. If reason requires that the act be done then surely one ought to do it, and the “ought” is based on all the reasons which apply to the case; it is a conclusive ought. But this entails that failing to perform the act is failing to do what one ought (conclusively) to do, so why isn't it blameworthy not to perform a supererogatory act? (Raz 1975a, p.164)

My first comment regarding his manner of framing the problem of supererogation is that the supererogatory action is called “praiseworthy” instead of “good”. This changes some nuances, but it does not change the connection with reasons: being praiseworthy guarantees having reason for doing the action, in the same way being morally good usually guarantees having reasons for doing it.⁹⁶ My second comment is that the general commonsensical question “How can one explain the permission to omit something morally good?” is replaced with the more theoretically laden question “why isn't it blameworthy not to perform a supererogatory act?”. Also, all reference to “duty” or going beyond a certain threshold is missing, as the normative force of “what one ought to do” is represented simply by having reasons for doing.

Before analyzing Raz's solution to the problem, I think one of his remarks is interesting to note. Raz takes into account the possibility of an anti-supererogationist reply, namely to say that not doing the prescribed praiseworthy action is, actually, blameworthy. In other words, he ponders the kind of answer that challenges the question: there is no explanation for the agent's being allowed to omit the good deed

⁹⁶ Raz (1975a, p.166) explicitly admits that he is working under the following assumption: “The fundamental assumption on which my argument is based is that if p is valuable then everyone who can do so has a reason to do whatever will make the occurrence of p more likely.”

because the agent is not, *actually*, allowed to omit the good deed. Raz's reply is, I think, quite illuminating for the kind of problem he is taking himself to pursue. This is because his answer shows that he does not take himself to be involved in a problem regarding the *existence* of supererogatory action, but its logical possibility:

One way out of the problem is to concede that it is blameworthy not to perform such acts and to maintain that there are no supererogatory acts. This answer may be correct, but even if correct it fails to solve the problem. Even if those who believe that there are supererogatory acts are mistaken, they are not incoherent. Those who criticize their views are exposing a moral mistake, not a logical inconsistency. How, therefore, is supererogation to be explained? (Raz 1975a, p.164)

Therefore, I think Raz takes very seriously the first part of his question about supererogation, namely the part that asks 'how can one explain (...)'. What he means by that phrase turns out to be something like 'can both claims about supererogation (i.e. that one can be both praiseworthy for doing and not blameworthy for not doing the same good action) be made in a theoretically coherent picture?' rather than 'is it true that one can be both praiseworthy for doing and not blameworthy for not doing the same good action?'. By contrast, Nagel took himself to solve a problem about the reality⁹⁷ of our ethical lives when he was trying to solve his version of the problem of supererogation. This is not to say that Nagel was a fully convinced supererogationist. That would be a bit simplistic. Of course, Nagel takes himself to have a theory that is supported by the existence of what he calls “supererogation”. However, I have argued that the supererogation he believes in is quite different from the one invoked by common sense.

⁹⁷ Nagel did not express doubt about the existence of supererogation, i.e. he took it to be a well-established thesis of common sense that there are deeds that one can rightfully call “supererogatory”.

The way Raz thinks the two claims about supererogation (i.e. praiseworthiness for action, and lack of blameworthiness for non-action) could be both coherently entertained (i.e. his solution to his problem of supererogation) is also a step towards making a more general point in his theory about conflict of reasons. The reason why he is interested in supererogation is because it shows that his theory of second-order reasons is useful, namely it is useful in solving the problem of supererogation. Let me explain how second-order reasons and supererogation are connected for Raz. He thinks reasons can come into conflict in more than one way. The usual way of thinking about conflicting reasons is in terms of opposing weights:

According to our intuitive conception of practical conflicts such conflicts are to be resolved by assessing the relative strength or weight of the conflicting reasons and determining what ought to be done on the balance of reasons. To put it another way, one ought always to do whatever one has a conclusive reason for doing. Or, which is another way of saying the same, one ought always to act on the balance of reasons. (Raz 1975b, p.36)

However, he thinks there is at least one other way in which reasons may be in conflict. This is because, Raz thinks, one can have not only *reasons for* or *reasons against* doing a certain action, but also *reasons about other reasons* (to take them into account or not). The latter are second-order reasons and the former are first-order reasons: “A second-order reason is any reason to act for a reason or to refrain from acting for a reason.”⁹⁸ In tackling supererogation, one will be interested only in the second kind of second-order reasons, namely in *reasons to refrain from acting for a reason*, named by Raz *exclusionary reasons*: “An exclusionary reason is a second-

⁹⁸ Raz (1975b, p. 39).

order reason to refrain from acting for some reason.”⁹⁹ Raz gives the following example of exclusionary reason: let us suppose that an agent is has to make a decision about an investment. However, there is a narrow window of opportunity and a decision has to be made in a couple of hours. The agent is tired, emotionally drained, after a hard day. Even though she can take into consideration reasons for and reasons against, she does not trust her judgment right now. Raz claims that what happens when the agent says that she cannot trust herself to make a correct judgment in the circumstances, is, actually, that the agent gives reasons for disregarding other reasons: “She claims to have a reason for not acting on the balance of reasons.”¹⁰⁰

The notion of an exclusionary reason is not straightforwardly helpful for the problem of supererogation. Supererogation, by allowing the agent to omit the action, seems to entitle the agent not to act on the balance of reasons (which balance of reasons is in favor of the action, because it is a praiseworthy action). However, if the reason why the praiseworthy action was omitted will be seen as an exclusionary reason, then the agent would be *required* to omit the action¹⁰¹ (i.e. it would be wrong to do the action). What is needed, then, is not an exclusionary *reason*, but an exclusionary *permission*:

How can one be permitted to refrain from action which is required by reason? The solution is to be found in the notion of an exclusionary permission. The permission to refrain from performing a supererogatory act is an exclusionary permission, a permission not to act on certain reasons. An act is a supererogatory act only if it is an act

⁹⁹ Raz (1975b, p.39).

¹⁰⁰ Raz (1975b, p.38).

¹⁰¹ Raz (1975b, p.40) has a principle according to which the second-order reasons always win if they are about first-order reasons that are conclusive: “Such conflicts are resolved not by the strength of the competing reasons but by a general principle of practical reasoning which determines that exclusionary reasons always prevail, when in conflict with first-order reasons.”

which on the balance of reasons one ought to do and yet one is permitted not to act on the balance of reasons. (Raz 1975a, p.165)

An exclusionary permission¹⁰² is also a second-order entity taking first-order reasons as its objects. An exclusionary permission cannot *require* the agent to omit an action (in the way an exclusionary reason does), it can merely *give permission* to omit the action:

Exclusionary permissions differ from exclusionary reasons in that they do not entail that one ought to disregard the excluded reasons. They merely entitle one to do so. I act against reason if I do not disregard reasons excluded by an exclusionary reason, whereas I am not acting against reason when I act on reasons which I am merely permitted to disregard. (Raz 1975b, p.90)

An example of exclusionary permission (other than supererogatory action) is, according to Raz, your giving someone permission to harm you.¹⁰³ As there are first-order reasons against harming, the exclusionary permission allows an agent to go against those reasons, but it cannot require the agent to do so.

Raz addresses the worry that exclusionary permissions might have a paradoxical appearance, but he thinks there is no paradox as long as we keep in mind that there are different levels to be taken into consideration:

That reason entitles one to act against the balance of reasons may be thought paradoxical. The appearance of paradox disappears if it is

¹⁰² Raz (1975b, p.90) discusses other two kinds of permissions and the difference between them: "There is yet a third way of justifying permissions. I am permitted to perform an act despite the fact that there are conclusive reasons for not performing it if I *may* disregard those reasons. A permission of this kind differs from a weak permission which is based on the absence of conclusive reasons for not performing the act. It also differs from a permission based on an exclusionary reason in that I do not have a reason requiring that I shall disregard conflicting reasons. I merely may do so. I shall call such permissions exclusionary permissions."

¹⁰³ Raz does not discuss if the supererogation case there is a difference between giving yourself an exclusionary permission and giving someone else an exclusionary permission. I think it would be an interesting discussion.

realized that to assert that an exclusionary permission is valid is not to assert that unreasonableness is reasonable, it is to assert that reasons of one level entitle disregarding reasons of another level. (Raz 1975a, p.167)

How, then, is the idea of an exclusionary permission going to solve the problem of supererogation as seen by Raz? To recall, a solution had to be able to put together coherently both the praiseworthiness of doing the heroic act and the lack of blame for not doing the heroic act. It is clear how exclusionary permission explains the lack of blame in case of inaction: the agent is entitled by this second-order permission¹⁰⁴ to go against the balance of first-order reasons. However, Raz claims, the reason to act heroically is not overridden by the exclusionary permission. This is because the relation between first-order reasons and second-order exclusionary entities is a conflict, but a different kind of conflict than the one in which the sides were weighed to see which one wins.¹⁰⁵ Therefore, the first-order reasons are not overridden by the weight of another reason, and so they still produce their effects by making the action praiseworthy (even if it is not required anymore):

It follows that the view we are examining presupposes that there are at least two ways in which reasons can be compared and conflicts between them resolved. If they belong to the same level they can be compared in point of strength. The stronger reason should prevail, the weaker should give way. But reasons may belong to different and independent levels of considerations (whether because they derive from different ultimate values or on other grounds). In this case their strength cannot be compared, but the different levels may still be

¹⁰⁴ Raz (1975b, p.94): “Since exclusionary permissions are required to explain the nature of supererogation there can be no doubt of their importance to practical reasoning. (...) They are *second-order permissions* in the sense in which exclusionary reasons are second order reasons. They are permissions not to act on reasons, whereas exclusionary reasons are reasons for not acting on reasons. “

¹⁰⁵ Raz (1975b, p.40) gives the following principle P1. “The introduction of exclusionary reasons entails that there are two ways in which reasons can be defeated. They can be overridden by strictly conflicting reasons or excluded by exclusionary reasons.”

related to each other and one way in which they can be related is explicated by the use of the idea of exclusionary permissions. This means that the existence of reasons of one level entitles one not to act on reasons belonging to the other level. But the reasons of the other level are not overridden. (Raz 1975a, p.168)

Up to this point, Raz's proposal seems to make some logical room for supererogation, by showing that there are some theoretical devices capable of explaining how one might answer the problem of supererogation. He also has obtained, by discussing supererogation, a nice illustration for a technical device within his theory, namely for a theory of types of reasons and permissions. However, saying that one might have a second-order permission to not take into account some first-order reasons for an action is a rather technical and general description. Raz wants to convince his readers that his type theory is not just an *ad hoc* device, but is actually presupposed by various manners of reasoning about conflicting reasons. The way to do that is to show that solutions that have been commonly proposed to the problem of supererogation presuppose something very similar to exclusionary permissions. Supererogation, claims Raz quoting Bernard Williams (1973), is usually seen as a conflict between two kinds of values: the value of human autonomy and the value of human welfare. The usual solution is to say that the agent “is required to promote human welfare to the extent that this does not seriously interfere with the pursuit of his own fundamental goals” (Raz 1975a, p.167). The thought expressed by 'to the extent that this does not seriously interfere with the pursuit of his own fundamental goals' is the one that, according to Raz, “presupposes that some exclusionary permissions are valid”. Raz believes that our intuition that the two kinds of values in conflict are incommensurable is vindicated by the idea that the strength

of the two kinds of values (human welfare and autonomy of the agent) cannot be compared and therefore they cannot be placed in a balance of first-order reasons, competing against each other. However, they can be compared in another way, one that allows the agent to admit that following the reasons given by human welfare is praiseworthy, even if in certain circumstances it is not required:

On this view human welfare and the agent's personal goals, though they cannot be compared as to their strength, can be compared in some other way which entitles us to say that whereas, in specified circumstances, the pursuit of one's personal goals is permissible and legitimate the promotion of human welfare is praiseworthy. (Raz 1975a, p.168)

3.12. Raz's Solution to the Problem of Supererogation

A first instance of Raz's problem of supererogation appears to be quite similar to the commonsensical version: "How can one be permitted to refrain from an action which is required by reason?" However, it emerges later that Raz intended the problem to be more about "how can one explain that theorists may coherently claim that the agent can be permitted (...)." than about "how can one *actually* be permitted (...)". This is because one might deny that there are such moral permissions, and Raz would argue that simply assuming or even arguing for this position has not answered his question.

Nevertheless, his considerations about a different kind of conflict between first-order reasons and second-order reasons may transform the initial problem. The problem of supererogation in this context becomes 'how can one explain that first-

order reasons may keep their relevance (i.e. may not be overridden) while not establishing a requirement for action?' Raz's answer was to propose the concept of exclusionary permission.

I think it is also interesting to notice that Raz, on many occasions, uses the phrase “going against the balance of reasons” when he speaks about exclusionary permissions and exclusionary reasons. Properly speaking, what he describes is going against the balance of *first-order* reasons. The solution he finds is still taking reasons-standing-in-a-relation into account, even if it cannot be called a new “balance” of reasons. I believe that the construction of a relation between first-order and second-order reasons cannot properly be called a “balance” of reasons because the metaphor seems inappropriate this time. Between first and second order reasons there is no measuring of strength and therefore no work to be done by an imaginary balance.

Chapter Four: Supererogation, Obligation and Virtue

The previous chapters have offered a view into the diversity of problems usually disguised by the common label “the problem of supererogation”. Laying out the differences between various problems of supererogation reveals that they are not only about moral obligation, namely about explaining the non-obligatory status of certain good deeds. The problems of supererogation are varied because they reveal conflicts: for example, between supererogation and the obligation to obtain maximally good results (Slote) or supererogation and the impersonal aspect of morality (Scheffler) or, again, supererogation and the exclusively agent-neutral aspect of morality (Dancy).

This enumeration does not exhaust all the aspects discussed in the previous chapters, but it nevertheless shows a remarkable variety of standpoints from which supererogation is thought to be problematic for a moral theory. However, my claim is that this variety does not amount to mere randomness, i.e. I believe that the aspects raised by the various problems of supererogation are, most of them, related to each other. To be more specific, I believe that most aspects being disturbed by supererogation (such as aiming at impersonal good, being obligatory, being agent-neutral and impartial), may be said to form a certain traditional view about morality in general. The problems of supererogation appear when several traits characteristic to supererogation come into conflict with several traits of this particular view about morality. These traits are not all independent, they have conceptual links allowing them to form a certain general view. For example, if morality is thought to be centered upon *obligations*, and obligations are thought to be established by means of general rules, and these general rules are taken to resemble laws, then this may very well result in an *impersonal* character of moral rules (following the model of the law).

This is why I have claimed, in the beginning, that the problems of supererogation are *different, but related* problems, or a “family” of problems.

In short, the main point to be established in this chapter is that taking into account the difference between various problems of supererogation may say something about a common way of seeing morality, and may further lead to understanding why supererogation seems unwelcome in so many traditional moral frames and for so many theories. My answer is that supererogation is difficult to accommodate in many moral theories especially because those theories share specific presuppositions about what morality should look like. In the next section I am concerned with the details of my thesis and the possible objections to it.

4.1 The Main Points and Possible Objections

The core of my claim is that by taking into account the differences between various problems of supererogation, one will also be made aware of different aspects of one specific view about morality. This is a tenet that most likely needs to be detailed.

The remarkable proliferation of problems of supererogation comes, I believe, from various aspects of supererogation coming into conflict with some other aspects of a general view about morality. The view about morality I am only alluding to here (it will be described and discussed in more detail in the next sections) is one that has been described and criticized before, for instance by Peter Railton:

Morality may be conceived of as in essence selfless, impartial, impersonal. To act morally is to subordinate the self and all contingencies concerning the self's relations with others or the world to a set of imperatives binding on us solely as rational beings. We should be moral in this view, because it is ideally rational. (Railton 1984, p.165)

What Railton does not say in this quote is that what is most likely (even if not logically necessary) to be associated with selflessness, impartiality and impersonality is a law-like aspect of moral rules. This is why I will sometimes call this view the 'legalistic view of morality' or 'morality of law'. Usually, these traits of morality, seen in a legalistic manner, are tacitly presupposed, as this is a quite common view relying on a venerable philosophical tradition (as I will argue in the next sections). The problems raised by supererogation help bring these presuppositions to the surface because the problems are, actually, manners of questioning these presuppositions, one by one, from various angles. What the problems of supererogation bring to one's attention are oppositions, incompatibilities between traits presupposed by supererogation and traits presupposed by a legalistic view of morality. The difference between the various problems of supererogation come, in this view, from the differences between the different traits in conflict. However, as these traits are not unrelated, but form a coherent larger narrative, the problems also bear some resemblance to each other, forming what I have called a 'family' of problems. This is why, I think, they are different, but related problems.

Before delving into the details of each opposition and their connections, there are two objections that can be made to my claim from the start. The first is that it is implausible to say that such a specific, “parochial”¹⁰⁶ problem like supererogation can shed light on entire general conceptions about morality. After all, what could second order permissions (Raz) have in common with maximization and satisficing (Scheffler) or with an obligation to follow moral ideals (Pybus)? No easy-to-find answer seems to be within reach.

In spite of these considerations of heterogeneous interests, the second objection might be that the connection between supererogation and general theoretical frames is obvious: of course the problems of supererogation have something to say about morality in general – the problems differ depending upon the general theoretical frame in which they are formulated. The problem of supererogation in a utilitarian setting will be different from the problems of supererogation in

¹⁰⁶ Heyd (2015).

any other theoretical setting and it will bring forth some particular traits of that frame (e.g. maximization and its importance for the identity of the frame).

To answer the first objection one should point out how the particular concern of supererogation is, indeed, relevant to a general way of seeing morality as a legalistic morality. This is a task to be undertaken in most of the sections to follow. What I can say already is that the crux of the problem of supererogation is obligation (or lack thereof); when supererogation is pitted against a view of morality that takes obligation to be its *central* theme, then it is plausible to say that the attempt to integrate supererogation will have broad, general repercussions for the theoretical frame.

My answer to the second objection is that one way of differentiating between problems of supererogation (according to their being formulated in utilitarian or duty-based frameworks) does not preclude other ways of differentiating between those problems, namely it does not preclude a connection of those problems with other kinds of theoretical frameworks, ones that are not so visible and obvious. It is, indeed, quite obvious that problems of supererogation differ as theoretical frameworks are changed: I have pointed out not only how these frameworks shape supererogation differently, but also how the problem of supererogation, once formulated in terms of a framework, leads to changes within the framework (e.g. giving up maximization for satisficing). Therefore, it is true that it is somewhat obvious that considering the problem of supererogation may lead to considerations about entire theoretical frameworks. However, the theoretical framework brought to light by considering the various problems of supererogation is a different one, one that is harder to discern because its presuppositions are more general, tacit and not often discussed. Therefore, my tenet is about a different kind of generality or, maybe, a higher degree of generality. The view that morality is and should be legalistic may be supported in a utilitarian, deontological or reason-based frame, alike. Therefore, the relevant difference for my thesis is not the classical, well-known difference between these frames, but another one. I

have stated briefly that the general view about morality that I consider to be the origin of many conflicts involving supererogation is one in which moral rules resemble laws in several ways (peremptory in style and presupposing sanctions in case of disobedience, for example). I have called this way of seeing morality, 'morality of law'¹⁰⁷ and I intend to show that a morality of law of any persuasion (utilitarian or deontological) is, usually, an unfriendly environment for supererogation. One question that might be raised is what would be the alternative to this way of viewing morality. In short, one alternative way of seeing morality (to be presented at the end of this chapter) is what I have called, following MacIntyre, 'morality of virtue'. My position is that a certain view, that can be said to belong to virtue theory, might be a more welcoming environment for supererogation. In the last section I will defend this idea, that virtue ethics is better suited to explain supererogation, against recent counter-arguments by Heyd (2015).

4.2. Problems of Supererogation and What they May Bring to Light

In this section I will try to show how one may move from specific questions about supererogation to some general trait(s) presupposed by a certain picture of morality. I will start by recapitulating the journey through various problems to supererogation. After analyzing and commenting on them, I will argue that, not all, but a solid majority share a certain general view about morality.

The general form of the problem of supererogation is 'How can one explain that some type of action is morally good but not obligatory?'. Schematically this is 'How can one explain that something that is A, is not also B?' The presupposition here is that, normally or intuitively, *A should have been B*. In other words, A and B belong together usually but strangely enough, when it comes to supererogatory action, they are not united. Therefore, supererogation prevents some

¹⁰⁷ I am following MacIntyre here, with some reservations to be discussed.

association of traits of moral actions that, one assumes, is an association expected for other kinds of moral actions. From analyzing the table below one might extract these traits, i.e. the A and the B that usually are seen to go together.

Table1

	How can one explain that some morally good deeds	are not obligatory?
Urmson (1958)	How can one explain that some saintly and heroic deeds	are neither merely permitted nor obligatory?
Pybus (1982)	How can one explain that heroic and saintly actions	are something we ought to aspire to (even if the action cannot be demanded)?
Slote (2004)	How can one explain that maximization of the good	is merely morally permitted (non-obligatory)?
Dreier (2004)	How can one explain that choosing the lesser option, i.e. satisficing (in both moral and prudential domains)	is not irrational?
Scheffler (1994)	How can one explain that, in some circumstances, ignoring the demands of an impersonal scale of the good (i.e. establishing an agent-centered prerogative)	is morally permitted (non-obligatory)?
Zimmerman (1993)	How can one explain that, of two equally permitted moral actions,	one (the supererogatory) is better than the other?
Dancy (1993)	a) How can one explain that, as the value rises, the change in degree of evaluative property (i.e. the good) b) How can one explain that choosing not to follow agent- neutral reasons (and following agent-relative reasons)	is not attended by a change in degree of deontic property (i.e. the obligation)? is sometimes morally legitimate, i.e. permitted?
Nagel (1986)	How can one explain that morally excellent action involving harm for the agent	is not irrational, i.e. there are no decisive reasons against it?
Raz (1975)	How can one explain that some first order reasons may keep their relevance	while not establishing for the agent a requirement to act upon them?

The questions themselves are quite heterogeneous, but if one takes into consideration the answers and the discussions surrounding the problems, then a few themes emerge: moral obligation, its deontic force, and impersonal morality.

First, the majority of questions are about moral obligation or lack thereof in case of supererogation. One of Urmson's concerns is to argue in five steps that some saintly and heroic deeds cannot be deemed obligatory because this would mean that all heroic behavior can be demanded or "exacted like a debt". Pybus is arguing against his position by saying that all excellent moral deeds should be considered morally obligatory. However, according to her, moral obligation should not be conceived in a 'legalistic' manner or as something to be 'extracted like a debt'; rather, moral obligation should be considered a moral ideal towards which one should aspire.

Secondly, the utilitarian preoccupation with supererogation seems to revolve especially around the requirement of maximization (and understandably so). Usually, authors of consequentialist or utilitarian persuasion are willing to accept that optimization or maximization can be made non-obligatory (in order to accommodate commonsensical moral intuitions and supererogation). To recall, this would allow the harmonious integration of the concept of supererogation because it would make it possible to say, within the system, that excellent moral deeds (the ones with the best outcome and involving great sacrifices from the agent) are not obligatory. The requirement to maximize or to optimize would make these actions obligatory, which is a counter-intuitive result of consequentialist and utilitarian theories.

Zimmerman (1993), in his search to accommodate both maximization and supererogation, is an exception and a rare case. Dreier, Slote and Scheffler, on the other hand, accept that maximization (or optimization) can be made non obligatory, but in different manners. Slote (2004) argues straightforwardly in favor of accepting moral satisficing (doing less than the

best) as a rational choice and a good theoretical alternative to maximizing. However, Scheffler (1994) takes a another route to the same theoretical outcome (the desired theoretical outcome is the ability to integrate supererogation in a consequentialist system). He claims that one should be willing to allow for an agent-centered prerogative, a prerogative which allows the agent to ignore the impersonal scale measuring which deed will have the optimal effect.

More specifically, I believe that a plausible agent-centered prerogative would allow each agent to assign a certain proportionately greater weight to his own interest than to the interests of other people. It would then allow the agent to promote the non-optimal outcome of his choosing (...). (Scheffler 1982, p.20)

In other words, the agent is permitted, in certain circumstances, *to ignore the impersonal demands of morality* and to follow, instead, personal preferences or plans. This will result in being allowed to fail to optimize sometimes, but only because a *personal* way of assessment is being employed instead of following the impersonal one.

This is an example of a clash between an *impersonal* trait of morality and supererogation. It is not obvious why there should be a conflict between the concept of supererogation and the utilitarian/consequentialist way of measuring moral good impersonally. What, then, gives morality its impersonal trait in this context? Having an impersonal scale measuring the goodness of the outcome of an action, means that the measurement takes into account no specific point of view, no personal or partial perspective. It is not, therefore favorable or unfavorable to anyone, and it aims at something that can be considered good by as many as possible: an objective, overall good – the 'greater good'. The presupposition in this enterprise is that something impersonally evaluated is *objectively good*, i.e. *it is good for everyone*,¹⁰⁸ *overall*. A further presupposition (which will be discussed in the next section) is that the action with the best outcome for all is *rationally and morally required*: there are simply no considerations against it. One could have dismissed the obligation to act if the obligatory status of that act had been

¹⁰⁸ Of course, one has to count the sacrifice of the agent among the disadvantages, but when the outcome is sufficiently good, the sacrifice of one does not surpass the good obtained by many.

obtained from a subjective evaluation, from a certain *personal* point of view or would involve *partiality*. From these rather banal considerations, a certain picture of morality starts to emerge: morality and its obligations (which are central to it) rest on *rational, objective, impersonal and impartial* considerations. Moreover, concerns about the sacrifice demanded from the agent are not prominent, even if they have been included in the measurement of the resulting objective good. Especially in a consequentialist/utilitarian frame, what is deemed to be objectively and impersonally best, is also obligatory.

In this context, it is clear that extraordinary good moral acts, producing optimal results *for all*, are deemed obligatory, against the main tenet of supererogationism that extraordinary good deeds cannot be required – this is where the conflict between supererogation and impersonal evaluations of the good intervenes. The implication from objectively best to 'ought to be done' is at the core of the conflict.

Another trait of morality that can be extracted from the table of problems of supererogation is *demandingness*. This view, that morality is difficult to achieve, and its obligations are demanding, brings to light the worry about the *moral agent's burden*. Among consequentialists, Scheffler is particularly interested in this problem.

Not giving enough weight to the agent's losses, according to Scheffler, has larger implications for the relation between a system of morality and the agent acting morally. These implications go beyond the difficulties presented by the case of supererogation. In other words, not being able to accommodate intuitions about supererogation is just one of the problems created by the impersonal way of measuring the good. Scheffler (1982) proposes his agent-centered prerogative as a response to a famous critique by Bernard Williams, and not necessarily as a solution to the problem of supererogation (even if it turns out to be that, too). Williams claimed that making morality impersonal (for both the utilitarian and Kantian viewpoints) leads to an alienation of the moral agents from their own life projects. As the objective, impersonal

good is considered not only good for everyone but also *obligatory*, the result is that the agent is required to evaluate all personal projects by reference to the impersonal scale measuring the capacity to increase the overall goodness or happiness (in the case of utilitarianism). This means that the agent will be morally allowed to put into practice only those projects that align with the general good – which will leave very little room for the personal side of one's life. This is a different objection from the objection that the maximization requirement demands sacrifices that would involve serious harm for the agent and therefore the demand is unreasonable (i.e. the demandingness objection). Scheffler's agent-centered prerogative is intended as a response to both objections, though.

I will not discuss here how much success Scheffler has in responding to these objections. What is relevant here is the image of morality one can extract from this kind of response to this problem of supererogation: morality emerges as *obligation-centered* (the main problem is what is obligatory and what is not), *rational*, *impersonal*, *impartial* and quite *difficult* for an ordinary agent to comply with (see the demandingness objection and Williams' objection regarding alienation from one's own plans).

There is one other coordinate to be added to this image resulting from the overall evaluation of the table above. It is something not entirely new, as it regards the nature of moral obligation. But it useful because it adds to the above picture of morality some details about how obligation is supposed to work.

A third trait of morality that can be extracted from the overall picture of problems of supererogation regards *reasons and agent-neutrality*.

Dreier (2004), in his analysis of satisficing, states that a) if some action X is deemed objectively better than any other action available, then this amounts to having reason to choose X; and b) once we have reason to do X, failing to do X *needs some explanation or justification on pain of irrationality*. This tenet, that one cannot simply fail to follow an objective reason

without some explanation or without having an opposing reason is one widely shared in reason-based theories. The underlying presupposition for this tenet is that reasons, even if they do not amount to full-scale obligations or 'oughts', have, nevertheless, something called 'deontic force'. I think Little (2013, p.112) explains it best:

On a certain classical picture, reasons, by their very nature, are deontic. If one faces a reason in favour of an action, then one would be *wrong* not to follow its lead, absent sufficient counter-veiling justification. Depending on the type of reason, the wrong in question might be the wrong of immorality, or imprudence, or just the generic wrong of practical reason (what I will refer to in this chapter as 'irrationality' or 'impermissibility'). (...) Reasons for actions are normative entities inherently on their way toward being all things considered deontic oughts. Call this the deontic view of reasons.

This view of reasons – as deontic forces that have to be first countered by other equal 'forces' in order for one to be allowed to ignore them with no consequences – explains the way in which Raz frames his version of the problem of supererogation:

If doing a supererogatory act is praiseworthy, there must be reasons for doing it, and the reasons must outweigh any conflicting reasons for not doing it. But if there are conclusive reasons for performing the act then not to perform it is to act against the balance of reasons. If reason requires that the act be done then surely one ought to do it, and the "ought" is based on all the reasons which apply to the case; (Raz 1975a, p.164)

This is another version of the connection discussed earlier between good and obligation, where obligation is replaced by the deontic force of reasons. The problem is that, once it is established that there is a conclusive reason for doing something, then it is hard to see how its omission can be justified. Nagel (1986) does not disagree with this position, but he chooses to focus his version of the problem on another aspect. For him it is important to defend the *rationality* of moral action, namely to find theoretical means for supporting the thesis that the agent who acted morally and suffered great losses as a result, still acted *rationally* – even if the well-being of the agent was greatly affected. Nagel is also a very good example for the thesis that morality should be as *objective, impersonal and impartial* as possible: this is his great quest in

establishing a “view from nowhere” in ethics, even if he accepts that this demand has to be mitigated by obvious human limitations.

From the positions surveyed here, only Dancy (2004) is willing to go against the dogma of the deontic view of reasons and to admit that there might be other kinds of reasons, that are less “peremptory” in style. However, according to Dancy, this other kind of reasons (enticing reasons) cannot be used to give a solution to the problem of supererogation. His way of solving the problem will consist in allowing agent-relative reasons to have moral weight, one that can be opposed to agent-neutral reasons counting in favor of heroic action. It is worth underlying that allowing agent-relative reasons to have moral weight means that the agent's viewpoint is taken into account in a different way. Usually, the loss of the agent was counted as a *prudential* reason to not do a risky but beneficial action, not as a *moral* reason. Dancy allows the cost calculated “for me” (as opposed to “anyone”) to count as a moral consideration in establishing the balance of reasons. I think that his solution goes against a longstanding tradition for which only agent-neutral reasons can have moral weight – which is another way of saying that, traditionally, morality should avoid any particular point of view or consideration, i.e. it should be *impersonal* and *impartial*.

To conclude, most of the problems of supererogation discussed have an underlying presupposition about some traits that should be conjoined, but in the case of supererogation they are not. The general form of the questions listed is 'How can one explain that something that is A is not also B?' – how can one explain that some saintly, heroic, morally excellent deeds are not also obligatory, rational, impartially demanded? Of course, A stands here for a description of the supererogatory act, i.e. of a highly moral act. However, acts of the type A are also supposed to be included in the larger class of morally good acts and, as a consequence, they should also be B, i.e. whatever trait morally good acts have associated with them generally.

The most common tacit reasoning behind these question seems to be that morally good

deeds are usually obligatory, so it is surprising that supererogatory acts, which are morally good, cannot be deemed obligatory. For example, saintly or heroic deeds are morally good deeds and therefore they should be considered obligatory (they cannot be wrong or merely permitted as they are not indifferent to morality), where “obligation” is taken to have a “legalistic” meaning, as something that can be demanded.¹⁰⁹

Summarizing the analysis undertaken in this section, these are the traits that are supposed to be in close association (but fail to be in the case of supererogation):

- a) being a *morally excellent act* and being an *obligatory act* in a strong, *legalistic* sense.
- b) being assessed as morally best act in an *impersonal, impartial* and *objective* manner (i.e. being good for all, or *the greater good*) and being an *obligatory act* ; this will be also a *difficult or demanding* act for the agent (as the greater good usually implies *sacrifices* from the agent).
- c) being an action favored by *moral/agent-neutral* reasons and being something one *ought to do* unless there are other moral reasons going against the action favored in the first place.

The question is now if all these traits may form a coherent image of morality. Up until this point one seems to be met with the following narrative: if something is deemed to be morally good in an impersonal, impartial, objective manner (i.e. good in general, for many), then that should also be deemed obligatory even if it means to impose demanding (sometimes harmful) obligation upon individual moral agents. The worry had by the agent regarding his own losses does not have moral standing, but only prudential value, and so it is not deemed, from this perspective, to be a good enough reason for not imposing these obligations.

It is evident that there is a presupposed connection between being morally good and being obligatory or, at least, between being morally good and having a deontic force “pushing” in the direction of that good. This is a version of the 'good-ought tie-up' that was discussed in the first

¹⁰⁹ Of course, Urmson does not agree with this presupposition, as he is arguing against it. But the fact that he argues at length against it shows that it is not something to be dismissed lightly.

chapter. But the hypothesis that everything that is morally good has to be obligatory or at least *prima facie* obligatory, is not a very intuitive hypothesis. Where does this supposition come from, of the connection between good and obligation, and what is the connection between this 'good-ought tie-up' and all the other traits (being rational, objective, impersonal or impartial)?

4.3. Why Does the Good Have to be Followed by an Ought?

My hypothesis is that one source for the claim that any morally good deed should stand under an 'ought' is a commonsensical moral intuition. One other source is theoretical, and comes from a venerable tradition of theorizing about morality and its role in society. This section is dedicated to making explicit the first source, namely the intuition behind the claim (it will be followed by a section explaining the theoretical source, too).

That commonsensical moral intuition would be a source for this claim seems unlikely because the claim appears counter-intuitive. But it appears counter-intuitive, I believe, because in this particular formulation (all morally good deeds should be obligatory) makes evident an unintended consequence of the initial intuition, an intuition that would be much more appealing in another formulation.

The problem of supererogation itself, I believe, starts from equally strong intuitions pulling in different directions. On the one hand, it seems obvious that agents may go beyond duty in their moral endeavor: the doctor who rushes into a plagued city or the soldier sacrificing himself so that his fellow soldiers may live (to use Urmson's famous examples)¹¹⁰ are commonsensical obvious cases of agents going beyond duty.

¹¹⁰ In J.O. Urmson, "Saints and Heroes" .

On the other hand, it might seem that, in general, anyone who can help, *should* help others in a difficult situation. If this 'should' is translated into moral obligation, then each time one helps, one is simply fulfilling a moral obligation, something that is not, morally speaking, optional. I think this is the basic intuition behind the good-ought tie-up, namely that *one has a moral obligation to help others* in need because this is what constitutes moral good – one might say that this is what morality is about. The corollary of this thought is that one cannot invoke the inconveniences or the losses suffered by oneself in order to opt out of moral obligation: this is how morality works, by foregoing one's own interests to altruistically care about other people's interest. *If morality would make altruistic deeds merely permitted, then it would seem to be a declaration of indifference towards its own core: doing good to others, being altruistic* – this cannot be merely permitted as if it were indifferent for the moral agent and for the moral judgment whether one acts altruistically or not. Unfortunately, if one takes 'should' and 'obligation' in the sense discussed by Urmson, as meaning something that can be extracted “like a debt”, then this high-minded intuition about morality clashes with the other one, about 'going beyond the call of duty'. Of course, this is not a necessary move. The moral intuition I am discussing here has a vague enough form to be interpreted very differently. Nothing prevents a case-by-case interpretation instead of a rule of sweeping generality. Specifically, the moral intuition could be interpreted as saying that one should help someone in need whenever they have good reason to judge they are able to help and the need is, indeed, a serious one – which is a case-by case judgment depending on the agent. This is an interpretation that does not clash with the other intuition, about going beyond duty. However, this intuition can be interpreted in either two ways – as a sweeping obligation or as one that arises only in particular situation.

There are several conditions that need to be met by an interpretation in order to obtain the conflict between supererogation and the good-ought tie-up. First, it would have to be interpreted as a universal rule, without exceptions: *each* instance of altruistic behavior should count as obligatory. Moreover, it should *not depend* upon the agent to decide about

circumstances, ability to help, need to help, what counts as altruistic and so on. All these should be established impersonally and impartially such that no suspicion can be raised that someone might just find excuses not to act morally. In the same spirit, of impartiality and impersonality, the interpretation should not distinguish between obligations one might impose on oneself and obligations one might impose on others (similar to Slote's "self-other asymmetry").¹¹¹ In stating the intuition above, the statement is ambiguous between 'one should help' as something said by an agent to herself and 'one should help' as an expectation of everyone from everyone.

There might be other interpretations and other contributions needed to arrive at the good-ought tie-up, i.e. at the point of conflict. But I think that the shortest way to the conflict is to interpret "should" or "ought" in the sense of legalistic obligation (the universality of the rule, the impersonal and impartial aspect would follow automatically in this case). If one considers that the doctor had a moral *obligation* to help the plagued city and the soldier had a moral obligation to rescue the others, then the doctor and the soldier are not going "beyond and above duty", they are merely fulfilling their duties; which might be equally difficult and meritorious, but would not be supererogatory.

The conflict between these two commonsensical moral intuitions becomes visible at this point: if each genuine need of help generates a moral obligation to help, then there is no action which could be properly called supererogatory (because any good deed is, roughly speaking, a kind of duty and therefore there is no 'going beyond duty'). Alternatively, if supererogation is possible, then some morally excellent deeds are above and beyond duty, i.e. they cannot be required. However, if their value is so great for us (and there is, I believe, little need to explain why a society might need kindness, generosity and other forms of going "beyond one's duty") then why are they not deemed obligatory?

¹¹¹ See Slote, M. (1984a) "Morality and Self-Other Asymmetry", *Journal of Philosophy* 81, pp. 179-192.

I think that the problem of supererogation reunites both parts of this dilemma, by accepting that there is supererogatory action while at the same time acknowledging the pull of the other side. In asking 'how can one explain that some morally excellent deeds cannot be morally required?', the problem presupposes the existence of supererogation while giving some credence to the opinion that if something is morally good then one *should* (as a matter of moral obligation) follow the good. What the problem of supererogation does is to ask for a justification for *denying* that every morally good deed falls under an obligation. Posing the problem, therefore, also means *that one should not take for granted that some morally good deeds do not fall under an obligation*. The problem contains, therefore, the claim that one needs to justify the permission offered by supererogation, to offer theoretical grounds for it.

There are attempts to dissolve the problem of supererogation by making a distinction between the “should” or “ought” of obligation and the “should” or “ought” of recommendation.¹¹² No matter how plausible one finds David Heyd's theorizing in this respect, I believe the problem persists under another formulation, namely: how could one explain that some morally excellent deeds are merely recommended, not required?

This is how, I think, the commonsensical formulation of the problem of supererogation appeals to both moral intuitions. That one ought to help whenever this is possible, is an intuition I have identified as one of the sources of the good-ought tie-up. Admittedly, the intuition alone, in its commonsensical formulation, is not enough to arrive at the good-ought tie-up. For this, I have claimed, 'ought' should be interpreted in a certain legalistic manner. But what would be the reasons in favor of such an interpretation?

I think a second source of the good-ought tie-up is needed to account for this interpretation, namely a theory of what morality should look like. The next section is dedicated

¹¹² For distinguishing between different meanings of “should” in the context of this problem, see David Heyd's article on supererogation in the *Stanford Encyclopedia of Philosophy*.

to finding out the reasons and the presuppositions behind the choice of a strong sense, a legalistic sense for moral obligation.

4.4. Law-like Moral Obligation

The question 'how can one explain that some morally excellent deeds are not required?' presupposes some degree of puzzlement or at least of skepticism about the idea that not all morally good deeds can be required. In other words, the puzzle seems to derive from joining excellent moral deeds with the permission to forgo them. If the permission is puzzling, then the presupposition seems to be that the non-puzzling state would be the one in which we are *not* allowed to omit any morally excellent deeds. Therefore, the obligation to act in a morally excellent manner would be the expected state.

Of course, this is not a plausible psychological or moral reality: to be required or morally obligated to give one's life to save others, or to risk one's health to cure other people seems an unreasonable demand. After all, the most intuitive answer to the question 'how can one explain that we are not obligated to do some morally excellent actions?' is that some of these actions involve great sacrifices from the agent.

This answer is not, in itself, a solution of the puzzle. First, because the question can be reiterated in search for a theoretical approach (i.e. 'how can one theoretically ground the intuition that great sacrifices cannot be demanded?') and second, because it is not always the case that good deeds that cannot be demanded involve great sacrifices.¹¹³ For example, a small token of appreciation cannot be demanded but it is not very costly for the agent – how could one explain that these are not to be demanded, as the heavy burden imposed on the agent cannot be a justification here.

¹¹³ For a lengthy discussion of cases where supererogatory action does not involve much sacrifice from the agent, see Horgan and Timmons (2010).

Theoretically justified or not, it is a strong moral intuition (the one that mainly supports the plausibility of supererogation as a viable concept) that at least some morally excellent and admirable actions cannot and *should* not be demanded from anyone. This is why act-consequentialist theories like Scheffler's (1994) and Slote's (1984) take into account supererogation when they want to reconcile their theories with commonsensical moral intuitions.

Therefore, if one agrees that this is a strong moral intuition (that some good actions cannot be required), then the question is: how can one arrive at the impression that the opposite might be justified as well? In other words, *how does one get to the idea that all morally good deeds should be required?*¹¹⁴

My answer to this question connects the *demandingness* of legalistic morality with its *obligatory* aspect. It is, indeed, counter-intuitive to see as obligatory every good deed - it would be too difficult to fulfill this obligation, and this conception of morality would be too demanding. My hypothesis about how one could reach the idea that any good action should stay under an "ought" is the following: *it is exactly because the claims of morality are seen as demanding and difficult to comply with, that they all have to be made obligatory*. The worry hidden here is that if some good deeds would not be seen as obligatory, i.e. as standing under an "ought", then no one would do those good deeds,¹¹⁵ and they would not be taken seriously (or maybe they would be seen as a matter of indifference).

This hypothesis, that moral good has to stay under an 'ought' in order to be translated into action betrays a supposed reluctance on the part of the moral agent to follow the good. This, in

¹¹⁴ One may find a partial explanation in John Rawls, *A Theory of Justice*, namely an explanation for why utilitarian theories might arrive at this idea. His theory is that utilitarian-inclined theories will extrapolate from what is rational for one individual to what is rational for a group and so the losses suffered by an agent in maximizing the good are similar to the losses that any one agent is willing to suffer for a long term gain: "We may impose a sacrifice on ourselves now for the sake of a greater advantage later. A person quite properly acts, at least when others are not affected, to achieve his own greatest good, to advance his rational ends as far as possible. Now why should not a society act on precisely the same principle applied to a group and therefore regard that which is rational for one man as right for an association of men?" (Rawls, 1971, p.23)

¹¹⁵ Gregory Mellema (1991) in *Beyond the Call of Duty* describes the historical dispute between the Protestant and the Catholic Church regarding supererogation, where the protestant position was that if the teachings of the Gospel are mere "counsels", not obligations, then nobody would be under an obligation to follow them, so nobody would take them seriously.

its turn, is a plausible supposition if one sees the *moral good and the personal good as being usually at odds* with each other. In other words, the picture makes sense if one accepts that, in general, being moral involves (smaller or larger) sacrifices from the agent. Usually (even if not every time) the higher the moral value, the larger the sacrifice (and therefore an explicable reluctance on the part of the agent). This is what I call the *zero-sum hypothesis* between the moral and the personal domain, namely the idea *that whatever resources the individual spends on his or her personal/ private side of life are somehow "lost" for society*, generally speaking (one may allow exceptions); and the other way around: whatever I do for the common good or for others will come at a cost, will be a loss or will consist in neglecting my own interest.

Even if today these considerations seem obvious (as any example about donating to charity tends to illustrate) this has not always been the case. The next section is a brief historical detour for finding an alternative view.

4.5. Looking back: Virtue, Law and Morality

I will start from one astute observation made by MacIntyre about an important change in a venerable and important philosophical frame, namely how *morality of law* comes to replace the *morality of virtue*.

MacIntyre describes in *After Virtue* (especially Chapter 13) how gradually, over centuries, a certain picture of society and morality changes in Western Europe: it used to be the case that to follow the good meant to attend to one's virtues (moral and intellectual). This activity was not in opposition or contrast with attending to one's city or society or public affairs. And this was the *morality of virtue*.

However, this harmonious coexistence between what is good for the individual and what is good for the city was made possible by a common, shared conception of the good, according to MacIntyre. Slowly, as the belief in a common good disintegrates, the harmony disappears. As a result, attending to one's improvement as a person does not follow the same direction as the public good (the goods have become divergent) and so attending to one's improvement becomes a private affair unless it is explicitly and on purpose subsumed to the public good (i.e. virtues of loyalty, patience, charity, chivalry – all deemed important in the new picture).

There would be no genuine shared common good; the only goods would be the goods of individuals. And the pursuit of *any* private good [including virtues], being often and necessarily in these circumstances liable to clash with the goods of others, would appear to be at odds with the requirements of the moral law. Hence if I adhere to the law, I will have to suppress the private self. (MacIntyre 1981, p.170)

The morality of virtue is replaced with the following moral view: to follow the good is to serve *not* one's private purposes (as cultivating virtues comes to be seen), but the public order, the *law*.¹¹⁶ As a consequence, the public sphere and the private sphere are seen, by hypothesis, in opposition; virtue, if not an instrument for obeying the law, is relegated to the less important domain of private affairs (and so are the arts, games and sciences, notes MacIntyre).

The individual good pursued outside a frame in which the majority agreed about what the good is, was very likely to bring about conflict: someone's good would obtain at the expense of someone else's good. As a consequence, the law comes to be understood as placing a limit on individual enterprises, and thus defending society from collapse or violent disintegration. The law becomes in this way the new "common good", but a negative one this time, namely it only stops and prevents, it does not lead towards anything else: "The point of the law cannot be the achievement of some good beyond the law, for there now appears to be no such good."

¹¹⁶ MacIntyre tends to see this change of framework as regrettable. I do not share this attitude, as I believe both frames have their own advantages and disadvantages. This discussion is interesting and complicated and unfortunately cannot be undertaken here.

This way of thinking, according to MacIntyre, reaches its full bloom in the 17th and 18th century, when morality is seen as the instrument *par excellence* for curbing what was thought to be natural human selfishness. Broadly speaking, morality¹¹⁷ and law are altruistic, they take into account the others, showing equal concern for their interests; on the other hand, the private self led by passions, is more than often guilty of selfishness, a tendency that must be restrained if society is to endure.

It was in the 17th and 18th centuries that morality came to be generally understood as offering a solution to the problem posed by human egoism and that the content of morality came largely to be equated with altruism. For it was in that same period that men came to be thought of as in some dangerous measure egoistic by nature; *and it is only once we think of mankind as by nature dangerously egoistic that altruism becomes at once socially necessary and yet apparently impossible and, if and when it occurs, inexplicable.*¹¹⁸ (MacIntyre 1981, p.229)

One may notice that what MacIntyre says about altruism in this quotation holds, roughly speaking, also about supererogation: if man is thought to be egoistic by nature¹¹⁹, then doing good deeds and going the extra mile are going to be rare and badly needed. Also, the rare occurrence of supererogatory deeds will be considered strange and inexplicable as it would go against the supposed natural egoism of the agent. For example, in Nagel's case there is a worry that supererogatory behavior, even if highly moral, might be seen as irrational because it clearly goes against the agent's well-being. For supererogation to become strange in this conceptual landscape, it is not even necessary to assume that "egoistic by nature" has negative connotations

¹¹⁷ A similar opposition, between the common good of morality and the private good of the self, seems to be underlying sometimes contemporary moral discourse, namely an opposition between personal (or private) aims and impersonal demands of the society. Of course there are qualifications to be added to this loose characterization: we are not speaking about whatever someone wants, but about something that one could reasonably want for oneself (there is a touch of impersonality here). Nevertheless, one may notice that a moral agent in general is, in contemporary philosophical discourse, more than often seen as having to choose between what he or she reasonably wants as a private pursuit and the demands of the society: what is good for all (or only for the rest?) is seen as conflicting with what is good for oneself. It is in this direction, I believe, that Bernard Williams (in "Persons, character and morality") was speaking of impersonal demands of morality *versus* personal project ("There can come a point at which it is quite unreasonable for a man to give up, in the name of the impartial good ordering of the world of moral agents, something which is a condition of his having any interest in being around in that world at all", p.14).

¹¹⁸ My italics.

¹¹⁹ For a very helpful picture of how conceptions about human nature can influence ethical theories see Paul Woodruff's (1991) "Virtue Ethics and the Appeal to Human Nature".

(even if MacIntyre seems to assume that it does); it can mean only that people are naturally mutually disinterested,¹²⁰ that they are not interested in each other's goals.

Obviously MacIntyre places this framework change in a certain historical time, and claims that morality of law and morality of virtue were two ways of seeing morality actually adopted by the public or by certain philosophers. For the purposes of this essay I do not need these presuppositions: it could be that his historical claims are not accurate. The idea followed here is just the idea of a conceptually viable and plausible distinction between these two ways of seeing morality.

4.6. Morality, Obligation and Supererogation

In short, MacIntyre's idea that I use is that our overall conception of morality changes as our conception about the good changes. Where does his theory about morality of law *versus* morality of virtue lead regarding supererogation and its problems?

When the common good of society is seen as distinguishable from, but nevertheless aligned with the personal good, morality is also seen as aligned with the agent's interest. In other words, morality is not something that might endanger her/his well-being. By contrast, when the common good of society is seen as diverging from the personal good, the overall image of morality changes. Morality, being on the side of the common good, will also be on the side of reason, of impartiality, of the public domain and its laws. It will also be other-oriented, impersonal (as a condition of impartiality) and rule-based. Because morality is seen as being at odds with personal interest, its rules will be also seen as difficult to follow, as they go against the supposed natural egoism of individuals. Because of its difficulty and perceived demandingness, morality will also be thought as a matter of obligation. As it is not likely that agents will usually

¹²⁰ Rawls makes this distinction in *A Theory of Justice* where he assumes explicitly that individuals behind the veil of ignorance are rational and egoists in this sense, of mutual disinterest.

act against their own interest, morality is thought to provide the needed external impetus towards doing the right thing, namely its cohort of obligations.

The resulting picture is one of two opposing camps: on the one hand, morality is on the side of reason, law, public domain and obligation; on the other hand, in the opposing camp are passions or emotions,¹²¹ freedoms, self-interest, private domain and permissions. Morality seen this way is impersonal, impartial (like the law) and other-oriented (i.e. directed towards the good of others or the common good, selfless). It is also based on obligations because its demands are seen as harsh and so they need to be somehow enforced (even if this is not identical with enforcing by law, it does resemble the law in its imperative or peremptory character).

This is not a new way of seeing morality, but a very old one, which has been often supported but also criticized. Bernard Williams (1976) famously argued that the impersonal approach of both Kantianism and Utilitarianism regarding morality ignores the importance of a personal perspective or a life plan for the agent. Peter Railton (1984) also argues that morality viewed as selfless, impartial and impersonal can alienate the agent “from one's personal commitments, from one's feelings or sentiments, from other people, or even from morality itself”.¹²²

Morality may be conceived of as in essence selfless, impartial, impersonal. To act morally is to subordinate the self and all contingencies concerning the self's relations with others or the world to a set of imperatives binding on us solely as rational beings. We should be moral in this view, because it is ideally rational. However, morality thus conceived seems bound to appear as alien in daily life. (Railton 1984, p.165)

Railton notices (and disapproves) not only the image of morality as impersonal and selfless, but also the philosophical presupposition that moral agents are, by default, egoists. Consequently, the concern for oneself is supposed to be automatic and thereby does not need

¹²¹ For example, Martha Nussbaum in “Emotions as Judgments of Value” offers an explanation for regarding emotions as suspect and inadequate for moral judgment. She criticizes two traditions in this respect: first, the one claiming that emotions have nothing to do with judgment (they are “blind” forces) and second, that “the judgments with which emotions are connected are all *false*”, p. 206.

¹²² Railton (1984, p.134).

explanation, while the concern for another is an odd kind of beast and it needs an investigation:

Distracted by the picture of a hypothetical, presocial individual, philosophers have found it very easy to assume, wrongly, that in the actual world concern for oneself and one's goals is quite automatic, needing no outside support, while a direct concern for others is inevitably problematic, needing some further rationale. (Railton 1984, p.168)

Therefore, the legalistic view of morality or “morality of law” is not something unfamiliar or unnoticed. The novelty of my observations consists mainly in connecting this view with difficulties regarding supererogation. However, the familiarity of the legalistic view of morality can sometimes obscure the peculiarity of some of its claims. Indeed, its theses are so widely shared, its reign so uncontested, that morality of law can pass sometimes for *morality tout court*. Williams himself, when criticizing this view about morality (in his “Morality, the Peculiar Institution”, 1997), takes himself to speak about *morality* – without any other qualifications – not about a certain view of morality, that might have alternatives.

That morality of law has presuppositions that are still widely shared as presuppositions of morality in general – will be the main point of the next section. In this section I want to point out the connection between morality of law and supererogation. To remind, morality of law cares to distinguish between two main domains: the one concerning the *general good*, where reason, law, morality and obligation reside; and the other one, concerning *private good*, where passion, freedom, self-interest and permissions are at home. Once this distinction has been made, supererogation will be a problematic concept. This is because, since it professes not belong to the domain of obligation, the division would have to relegate it to the other domain, namely to the private good, together with the rest of permissions. But that is also the domain of the self-interest and private good and supererogatory deeds are rarely working in the interest of the agent. Additionally, once it is relegated to the private good domain, supererogation seems to be excluded from the morality (by following the division itself). It would seem, therefore, that supererogation cannot belong properly to any of the two domains. The core of this problematic

status, I believe, is that morality of law sees *obligation* and *morality* as inseparable, while supererogation presupposes their separability. This is also the core of the problem of supererogation: *being puzzled about morally good deeds that are not obligatory is actually being puzzled about morality functioning without the compelling force of obligations*. And this is a puzzle only when the presupposition is that the core of morality *is, actually, obligation*. In other words, it is puzzling that some morally excellent deeds are not obligatory simply because, generally speaking, what is morally good is also seen as obligatory (or as something one should do).

This presupposition is not usually made explicit probably because there are powerful moral intuitions going against it. However, it can be expressed in ways in which it appears more plausible; saying, for example, “one should always choose to do the morally best available action” is not likely to be met with the same skepticism as “all morally good deeds are obligatory”. No matter how one chooses to explain this difference,¹²³ my main point about this presupposition of 'ought' following the good, is that the suppositions behind this idea are also powerful and originating in a venerable tradition. These suppositions are, first, that morality makes some difficult demands, that are difficult because they usually go against personal interest (what I have called the zero-sum game between the moral and the personal domain); second, that the cost imposed by these difficult demands of morality will most likely not be freely undertaken (as humans are by hypothesis egoists). Therefore, obligation is needed in order to make sure that morality will not be abandoned in favor of pursuing one's personal interest. The opposition between morality and personal interest makes sense, in its turn, under the hypothesis that the public good and the private good of the agent diverge.

It becomes clear, in this way, that the problem of supererogation has deep conceptual roots in a certain view about morality in general. This is not to say, however, that the whole

¹²³ One might choose to distinguish between various meanings of “should” or “obligation” in order to explain this difference. Alternatively, one could argue that “one should always choose” is ambiguous between what someone might ask from themselves and what someone might ask from someone else.

picture is a matter of the past, a historical peculiarity. I believe that the presuppositions underlying the problem of supererogation can be very often found in contemporary writings about morality or supererogation, as I hope to show in the next section.

4.7. Thomas Nagel, Susan Wolf¹²⁴ and a Common Presupposition

That all good deeds are morally required is an unfriendly supposition regarding supererogation. One might expect to find this supposition in theories of consequentialist and utilitarian persuasion or in explicit anti-supererogationist positions like Pybus (1982) and New (1974). And this would be a fulfilled expectation. Nevertheless, the suppositions discussed in the previous section are common enough to be found in contemporary texts that can be considered supererogation-friendly.

I have chosen Thomas Nagel and Susan Wolf¹²⁵ because they both want to allow for the existence of supererogation and, therefore, one would not expect to find in their theories various supererogation-unfriendly hypotheses (that make supererogation problematic). Moreover, I want to show that the presuppositions behind the idea that all good deeds are morally obligatory are so widely spread, so common and taken to be so basic, that they can be found even in theories that are rival theories. Nagel and Wolf differ widely in their attitudes towards the domain of morality. Nagel believes that the domain of moral requirements should be expanded to comprise actions that were formerly considered optional (for example, giving to charity). By contrast, Wolf thinks that expanding the demands of morality might not be wise, as it would be detrimental to the well-being of the agent.

¹²⁴ Neither Nagel nor Wolf treat supererogation as their main topic in the writings I am referring to. Supererogation is presented in each case as a relevant example that makes a certain point about a certain general theory they propose. However, the problems raised by their approach are underlined precisely by the case of supererogation.

¹²⁵ My points on Nagel are not entirely new. The observation that supererogation does not exactly fit within his general framework has been already made by Jonathan Dancy in *Moral Reasons*.

My aim is to show that, even though they seem to be at odds, Wolf and Nagel still share at least one important presupposition about the zero-sum game between morality and personal interest; and this presupposition makes supererogation difficult to accommodate within their theories despite their best intentions regarding the acceptance of supererogation.

Nagel's aim is a kind of compromise between "the objective and the subjective standpoint", namely that there should be a threshold for moral obligation (even though it is hard to say exactly where) because "it is unreasonable to expect people in general to sacrifice themselves and those to whom they have close personal ties to the general good".¹²⁶ For Nagel, the two sides do not exclude each other in terms of understanding: we can see the justification for "personal claims" even if we take the impersonal point of view. Therefore, from the impersonal point of view itself, the personal point of view has some justification, its own rational place.

In spite of this attempt at reconciliation, supererogation proves to be a difficult step. Supererogatory action is, for Nagel, a manifestation of the objective, impersonal standpoint whenever this standpoint is not limited by tolerance for human weakness:

(...) supererogatory virtue is *adherence to the claims of impersonal morality prior to their modification to accommodate the normal limitations of human nature*.¹²⁷

This modification takes the form of a relaxation of these requirements through tolerance, as it were, rather than the discovery of new moral reasons that outweigh the original impersonal ones. If they had been outweighed, then there would be reasons against the type of sacrifice that displays supererogatory virtue: it would be wrong. As things are, it is merely not required. And those who undertake it nonetheless are praiseworthy for submitting themselves to the true strength of reasons that they could not reasonably be required to follow strictly, given the mixed character of human motives. (Nagel 1986, p.204)

I believe this is a very clear statement of the opposition discussed: one has to make room for the personal viewpoint through *toleration* because otherwise the impersonal claims of morality

¹²⁶ Nagel (1986, p.202).

¹²⁷ My italics.

would drastically restrict the sphere of one's personal standpoint; we are excused from not acting in a supererogatory manner as the impersonal morality would have demanded.

And yet, according to Nagel, not even toleration and relaxation of demands can make the zero-sum relation disappear:

The result for the relation between morality and the good life is that some of the starker conflicts will be softened by these reductions of moral demands due to tolerance. But *the conflict will certainly not disappear*, and that brings us back to the original problem of whether, in such cases, morality will always have the net balance of reasons on its side, as against the good life. I believe that the self-limiting modifications of impersonal morality that let in supererogation make this more likely, though they do not guarantee it. (Nagel 1986, p.204)

Even if we "strike a bargain between our higher and lower selves in arriving at an acceptable morality",¹²⁸ as Nagel wants, this would be a compromise, not the proper way things should be.

On the opposing side, one energetic protest against the idea that morality takes precedence over all personal considerations comes from Susan Wolf (1982). She is clearly against the idea that we need *permissions* in order not to act in a supererogatory manner.

As she appears to oppose Nagel, one might think that her position is not a good illustration of the frame of mind that makes supererogation hard to explain. But I believe Susan Wolf still supports the idea that we have to make a difficult choice between the personal and the moral side of our life; it is just that she sides with the personal in trying to restrict the claims of the impersonal, altruistic domain (a domain that she, like Nagel, considers to be identical with the moral domain).

She provides several arguments according to which our image of a fulfilled happy life is quite far from those of a hero or a saint. The lives of heroes and moral saints, she claims, are impoverished lives from the point of view of well-being. Two cases are proposed to prove her

¹²⁸ Nagel (1986, p.202).

point: the case of the Loving Saint and the case of the Rational Saint. They both spend their time helping others, in altruistic endeavors, thereby accepting many sacrifices themselves. The difference comes from their attitudes: the Loving Saint is the one joyously sacrificing his time and resources, while for the Rational Saint the feeling of joy is not as present (even though the rational justification always is). Wolf's purpose in presenting these two types of moral saints is to convince us that the life of a moral saint is an impoverished life, no matter whether we look at the Loving or at the Rational Saint. Their life would be one in which they have sacrificed so much that there is no place for themselves anymore: for hobbies, friends, humor and so on. Her conclusion is that the life of a moral saint is not a desirable one:

In pointing out the regrettable features and the necessary absence of some desirable features in a moral saint, I have not meant to condemn the moral saint or the person who aspires to become one. Rather, I have meant to insist that the ideal of moral sainthood should not be held as a standard against which any other ideal must be judged or justified, and that the posture we take in response to the recognition that our lives are not as morally good as they might be need not be defensive. (Wolf 1982, pp.94-95)

Therefore, according to Wolf, we should not maintain that the moral ideal is the ideal to be followed completely and at any cost, because this leads to a mutilation of personal fulfillment.

No matter if one agrees or not with this recommendation, I think that the important assumption here is that one has to choose between the personal, non-moral, self-fulfilling life and the impersonal, self-eviscerated, moral life. But the supposition that morality sides mostly with the public/impersonal aspects of our lives and has no major role to claim in personal or private endeavors was the same zero-sum supposition that made supererogation difficult to explain.¹²⁹

¹²⁹ One possible objection to my interpretation is that Susan Wolf is not trying to give an account of supererogation in her article, therefore the difficulty does not arise for her. However, Wolf's central image is the moral saint, which is the paradigm case for supererogation.

4.8. Supererogation and the Theory of Virtue

To recall, the purpose of this research is not to offer a solution to 'the' problem of supererogation, since there is no single problem of supererogation. However, based on the previous chapters, I think I can venture some hypotheses about theoretical environments that would be favorable to supererogation, i.e. in which it would be more difficult for problems of supererogation to appear. I have already argued that certain theoretical environments are not welcoming for supererogation, they have difficulties in integrating the intuitions behind it in a theory about morality. For the remaining sections I would like to just point towards what could be regarded as a more welcoming theoretical environment for supererogation.

The considerations presented in the following sections are more akin to directions for future research rather than a fully developed position. However, I thought it might be important to sketch a way out from the entanglement of problems of supererogation.

I will start with a short reminder about what was it that made supererogation difficult to integrate, and then I will have some suggestions about how the obstacles can be removed. MacIntyre's labels, namely 'morality of law' and 'morality of virtue', will prove to be quite useful in suggesting the new direction, even if my conception overlaps only partially with his.

To recall, the morality of law was constituted by a complex view about humans, society and moral behavior. A short, rather unsophisticated version of the picture, can be given in the following steps:

1. One's own good diverges from the good of others.
2. Morality¹³⁰ ("of law") is essentially other-oriented and impartial, it is about doing good for others, and about preventing one from always acting in favor of one's own interest.

¹³⁰ I will here call "morality" what MacIntyre called "morality of law" (as opposed to "morality of virtue") because the hypothesis of divergent goods already places one in a context of morality of law seen as the only possible kind of morality.

3. Doing good for others comes at a cost for one's own good (as goods are divergent).
4. Therefore, doing the morally right thing comes at a cost.
5. The cost of morality will not be spontaneously or voluntarily undertaken (as people are thought to be naturally egoistic).
6. Therefore, in order for morality to be followed, a kind of law, a constraint, an obligation is needed. Therefore, acts of morality (i.e. of doing good for another) have to stay under an “ought” as they are an unwelcome burden on the individual.

There are several elements of this picture that might be said to be unfriendly towards supererogation, but I will take into discussion only two of them because I think they can lead to an alternative way of thinking about morality, one that can be more accommodating. These two elements of morality of law are its *focus on obligation and the zero sum game between one's own good and the public good*. I think that changing these two coordinates in ways that would make morality friendlier towards supererogation will lead to an account of morality broadly similar to the one espoused by the contemporary theories in *virtue ethics*.

Let us start, then, by analyzing how exactly morality's being obligation-centered interferes with supererogation. Broadly speaking, any idea or complex of ideas that inclines towards *expanding the domain of moral obligation* will present a threat for supererogation for obvious reasons. If all good, remarkable moral deeds are under an “ought”, then there is no room left for supererogation, no room for optional good deeds. But is the focus on obligation always going to lead to such an expansion of obligation? It would seem that there are theories that are centered upon obligation, but which *do* leave some room for supererogation. For example, Mill's conception about morality is obligation centered: the domain of Morality is built entirely around a strong legalistic notion of obligation; but he also speaks about the domain of Worthiness, where obligation is left behind. Therefore, it would seem that it is quite possible to have a

conception that is obligation-centered and also makes room for supererogation. Not only Mill, but also contemporary theories like Slote's or Scheffler's could be credited with the same kind of success. However, my point about morality being obligation-centered was not that no theoretical devices could be improvised to make room in such theories for supererogation. Of course, there are such theoretical devices being employed in several of the positions surveyed. My point regards, rather, a certain tendency of expansion of moral obligation, a tendency that must be curbed or limited by introducing agent-centered prerogatives or by expanding permission, for example. In other words, these theoretical devices are “pushing back” against the claims of dominance made by obligation in the moral domain. This idea, of the tendency of obligation to monopolize the domain of morality is noticed by Williams (1985):

There are actions (...) that are either more or less than obligations. They may be heroic or very fine actions, which go beyond what is obligatory or demanded. Or they may be actions that from an ethical point of view it would be agreeable or worthwhile or a good idea to do, without one's being required to do them. (...) How does the morality system deal with the considerations that seemingly do not yield obligations? One way in which the central, deontological, version of morality deals with them is to try to make as many as possible into obligations. (Williams 1985, p.49)

Williams thinks that “It is a mistake of morality to try to make everything into obligation. But the reasons for the mistake go deep.”¹³¹ He discusses several factors that might lead to such a tendency, such as reductionist attitudes (i.e. the tendency to reduce all important concepts to obligation) or, more prominently, the idea that “only an obligation can beat an obligation”¹³² and that any action deemed obligatory can find its justification in a general obligation. Combined, these two result in a tendency to judge merely permissible actions or indifferent actions as some special kind of obligation because otherwise they could not be justified: if there is a general obligation to help others, for example, then anything else one might do (i.e. attending to their own private business) can be morally approved *only* if it also has some moral weight, because

¹³¹ Williams (1985, p.50).

¹³² Williams (1985, p.51).

“only an obligation can beat an obligation”. The problem is that in obligation-centered theories this “moral weight” or justification can be given only by another, competing obligation. And therefore, the usage of the concept of obligation has the tendency to expand to the private domain:

But if we have accepted general and indeterminate obligations to further various moral objectives, (...) they will be waiting to provide work for idle hands, and the thought can gain a footing (I am not saying that it has to) that I could better be employed than in doing something I am under no obligation to do, and, if I could be, then I ought to be: I am under an obligation not to waste time in doing things I am under no obligation to do. At this stage, certainly, only an obligation can beat an obligation, and in order to do what I wanted to do, I shall need one of those fraudulent items, a *duty to myself*. If obligation is allowed to structure ethical thought, there are several natural ways in which it can come to dominate life altogether. (Williams 1985, p.52)

Williams' points depict the situation accurately, I believe, but they do not refer specifically to the relation between supererogation and moral obligation. I think that supererogation is not 'at home' in an obligation-centered approach to morality mainly because such an approach regards obligation as the *sole indicator of moral value*, or at least, the sole indicator of *important* moral value. In other words, if something is morally important (the supposition runs), then it will be codified in terms of obligation; if it is deemed to be not so important, then it can be left outside obligation and this sometimes has the consequence of placing it outside morality itself. Again, Mill's theory is a good example: actions that are deemed very useful, essential for the survival of society, are the ones deemed obligatory; the rest of moral action, evaluated as less central or important, or useful, can be left outside obligation. In this picture, if supererogation is admitted, it will be admitted as a curiosity, a peculiarity of commonsensical moral intuitions or as a “parochial”¹³³ moral phenomenon, albeit an admirable one. Therefore, even if the tendency to expand obligation is not followed, and some room is made for supererogation, this still means that the *important moral action takes place within the domain of obligation, not beyond it*. According to this point of view, what is morally important is

¹³³ Crisp (2013).

to attend to what one ought to do, to what we owe to others, to the moral common denominator, the moral rule or the moral law. What is beyond that, can be extraordinary in some sense, *but it is not the main preoccupation of morality*. In this light, Urmson's complaint that the category of morally permitted acts amalgamates very important moral deeds with indifferent moral deeds (e.g. sacrificing one's life is in the same category with buying apples), is quite perceptive. To put it in the usual metaphorical terms, an obligation-centered morality will deem important¹³⁴ the 'threshold' of duty; what goes beyond that might be sometimes morally significant but is something of an irregularity. What matters is to follow the rules and obligations. The extraordinary behavior of exceptional individuals (the hero, the saint) can be treated as an interesting moral quirk at best, not only because of the perceived difference in importance but also because the theory does not have the theoretical resources to speak about extraordinary good deeds, to differentiate or rank them or to ask what motivates them (other than obligation).

When trying to shift the accent from *the good that everyone should follow to the best someone could do (morally)*, a different view of morality emerges, one that is not based on obligation, but on moral *excellence*. In other words, its main preoccupation will be moral virtue. After all, it should not come as a surprise that a theory centered upon moral virtue would turn out to be more friendly towards supererogation than other theories: supererogatory deeds are usually defined as *excellent* moral deeds or as having higher value than what is usually required. And it is quite common for virtue ethics in general to be presented as an alternative to obligation-centered theories ("Certainly, it is characteristic of modern virtue ethics that it puts primary emphasis on aretaic or virtue-centered concepts rather than deontic or obligation-centered concepts.")¹³⁵ or as a critical alternative to legalistic ways of seeing morality.

Virtue ethicists have often accused the deontological conception of morality as

¹³⁴ This implication, between obligation and importance, was noticeable at several points in previous chapters: one presupposition in Urmson, for example, was that if something is not obligatory, then it is not important, it is indifferent to morality. The other way around for Mill: if the outcome of some type of behavior does not have a high enough utility, then it is not going to be deemed obligatory.

¹³⁵ Crisp and Slote (1997), Introduction to *Virtue Ethics*, p.3.

being too 'juridical' and overly preoccupied with justice, contracts, promises and rights. In other words, it is too minimal in its scope of norms and gives priority to prohibitions of causing harm over positive duties to do good. (Heyd 2015, p. 30)

Another, more vivid way of drawing the contrast between the two viewpoints on moral theory is in terms of the agent's motivation to act towards the good:

Virtues are sources of good behavior. Moral rules and laws set standards for doing right, but there is nothing about a rule that makes you feel like following it. In fact, there is something about many rules that makes most people feel like breaking them. According to virtue ethics, a good person is one who feels like doing what is right. (Woodruff 2001, p.12)

My second reason for maintaining that virtue ethics might be a more welcoming environment for supererogation is connected with the good life of the agent who is engaged in pursuing the good. This advantageous feature of virtue theory is often times called the *eudaimonist project*, i.e. “the idea that there should be some fundamental connection between what makes for a moral good life, on the one hand, and what makes a life worth living or intrinsically rewarding on the other.”¹³⁶

My claim is that the eudaimonist project, roughly speaking, is opposed to one crucial presupposition present in the picture of morality of law, namely the presupposition of the *zero-sum game between the agent's own good and the other's good*, the general good. In other words, the assumption is that, if one acts morally then this is a loss for the well-being of the agent and the other way around (an example is Susan Wolf's position discussed in a previous section). There is no overt opposition between the zero-sum game hypothesis and supererogation, but I think it is a rather intuitive idea that the more difficult moral actions appear to be, the likelier it will be for them to be made obligatory, because otherwise it becomes unrealistic to hope that agents will act, as a rule, against their own interest (that is, without a force to *oppose* their own interest – the deontic force of obligation). This is another foreseeable reason for which obligation

¹³⁶ Railton (2011, p. 305).

can be expanded to comprise the domain of all relevant moral actions. This expansion undermines the credibility and the possibility of a theoretically satisfactory account for supererogation: once the zero-sum hypothesis is in place, supererogatory action becomes socially necessary, apparently impossible, and when it occurs, inexplicable – to paraphrase MacIntyre.

Of course, there are obvious circumstances where no one can deny that acting morally takes its toll upon the well-being of the agent, but to say that this is *inevitably* the case or, at least, *generally* the case, is a completely different thing. Once the competition between the moral life and the good life is portrayed in these sweeping terms, it becomes a philosophical assumption, not a mere observation about moral life. This assumption is, I think, opposed to the idea that the moral life and the well-being of an agent can *enhance* each other:

Classical virtue theory held that asking how to live well, and asking how to act well, were closely allied questions. Indeed, these two were to be united in the life of the virtuous person: acting in the best way was best not only for the individual's family, friends, and *polis*, but for the individual herself. (Railton 2011, p.325)

This is not to deny that sometimes the harmony between the two is hard to achieve. In this respect, I agree with Railton's (2011) observations about the eudaimonist project as not being a very obvious position, especially in certain cultural environments:

Too often, it seems, virtuous action is its own punishment: standing up for what is right or doing what is best regularly collides with powerful interests bent on other purposes, and often is personally and socially 'inconvenient', to put it mildly. Much post-Enlightenment English-language moral and political thought has tended to take it for granted that the natural condition of most of us – even if we are fortunate enough to live together peaceably – is fundamentally competitive. And getting ahead – amassing wealth, power, recognition, or status – seems to have no intrinsic connection with moral worth. (Railton 2011, p.325)

However, Railton claims (in defending one of his favorite traits of virtue theory), that it is not too optimistic or far fetched to think that, at least in special circumstances, the eudaimonist project is a realistic claim. When the agent is in a social environment that she cares about (and I would like

to also point out the lack of impersonal context here), then it is more likely that the personal and the general good can be indeed brought together:

At least at this admittedly local level, doing well and living well can work to reinforce and sustain one another. And when genuinely painful self-sacrifice is called for, we may find ourselves with the enlarged psychological and interpersonal resources needed to do and sustain the harder thing. (Railton 2011, p.329)

To summarize my point about supererogation and virtue theory: I believe virtue theory has better premises to accommodate supererogation mainly because it is not obligation-based; this means that there is no danger of expanding moral obligation to deeds considered supererogatory in virtue theory, as morality is not essentially seen in terms of what is obligatory, what one ought to do. Another favorable premise of virtue theory is that it does not take the personal good and the general good to be at odds, locked in a zero sum game. Supererogation has better chances to be accounted for in these conditions because it will not appear as the strange occurrence of agents acting against their self-interest and natural inclinations *while no obligation is present*. One might notice also, when considering the zero-sum premise, the theoretical necessity of expanding obligation to cover all morally good deeds, especially ones presupposing sacrifice. For in a morality of law framework *there is no other explanation available for agents acting against their self-interest, but the force of the moral obligation*.

In the next section, when discussing objections to my claim, that theory of virtue is better equipped theoretically to account for supererogation, this idea will play an important role. I believe virtue theory appears to be in a difficult position to some authors because it appears to claim (among others) that agents act against their self-interest by sacrificing themselves *with no reason at all* (because the force of obligation has disappeared and, therefore, what else could lead someone to act heroically?). Of course, as I will show, this is not a fair accusation exactly because virtue theory is not committed to the zero-sum game hypothesis and has instead an approach corresponding to what Railton has called “the eudaimonistic project”.

4.9. The Case Against the Theory of Virtue

In the previous section I have argued for the thesis that virtue ethics might have better theoretical instruments to account for supererogation. Metaphorically speaking, supererogation would be more 'at home' in virtue ethics than in other classical theoretical frames (e.g. utilitarianism). My arguments are that virtue theory is not obligation-centered (and therefore it does not run the risk of expanding obligation, and it focuses on the moral behavior that goes beyond the threshold of duty), and it does not have the assumption of a zero-sum game between agent's morality and agent's well-being.

There are, obviously, other possible angles on the relation between virtue ethics and supererogation. Heyd (2015) argued that, contrary to my thesis, supererogation does not have a better chance of accommodation and theoretical justification within virtue ethics. He offers two main arguments for his position, the first one aiming to show that the concept of 'supererogation' is dependent upon the concept of 'duty', and a second one aimed at showing that one cannot have supererogation in an Aristotelian frame.

The first argument is based on a supposed correlativity of concepts: supererogation cannot be understood without a reference to a certain threshold of duty (beyond which the domain of supererogatory freedom is established):

(...) I want to suggest that supererogation and duty are mutually dependent. They are correlative concepts. Supererogation cannot be *conceptually* articulated without reference to duty and duty cannot be *normatively* justified without reference to what lies beyond duty. In that respect, supererogation is *essentially* a deontic concept or a phenomenon which is naturally accounted for in deontological theory. (Heyd 2015, p.31)

Heyd's aim here is to show that “a deontological framework is much superior to that of virtue ethics in both recognizing the special place of supererogation in moral behavior (...) and accounting for supererogatory action(...).”¹³⁷ According to Heyd, the superiority of the deontological framework consists in providing a clear threshold of duty, that would be necessary, to even be able to fully have a concept of supererogation. This is because “going beyond duty” indeed, analytically presupposes a duty to go beyond. However, I think Heyd equivocates here on the possible meanings of the word “duty”. Nothing in the commonsensical understanding of “going beyond duty” guarantees that the 'duty' in question is the one belonging to the deontological framework. In other words, supererogation might presuppose *some* concept of duty, but not *a certain* concept of duty, specifically the deontological one. The analytically presupposed threshold of supererogation may very well be a moral obligation imposed by social convention, convenience or utility. This is not the same kind of threshold that a deontological theory would want to designate by the term 'duty'. Mill and Urmson state clearly that, for them, social utility will determine where this threshold will be situated, and that it can vary with time and location.¹³⁸ Again, this is not what a Kantian will mean by 'duty'.

Maybe the point can be made clearer with an example: there is a law (in Texas) that certain seats in a public bus are reserved for senior citizens. There was a time when there was no such law, and giving your seat to a senior on the bus meant that one has done more than the law requires. But now, doing this benevolent gesture would be merely to comply with the law. Similarly, there might be moral rules that become rules from a certain point on, after a certain society has decided to add something to its list of moral obligations. Before, it was supererogatory to act according to the rule, but now it is not. A new threshold for supererogation was imposed. But surely this is not what the threshold of absolute 'duty' looks like in a deontological framework. In short, what I want to say is that other theories might have their own way of imposing a certain threshold for supererogation to go beyond. It does not have to be 'duty'

¹³⁷ Heyd (2015, p.26).

¹³⁸ I owe to Paul Woodruff the observation that the threshold could also be set by the concept of *human* virtue.

in the deontological sense. And, I think, virtue theory has its own means of accounting for such thresholds – more details about how this might be done will be provided when answering to the next argument.

Heyd's second argument is a bit more convoluted. He claims that virtue ethics cannot have the same partition that supererogation has, between acts that are excellent and go beyond what is required and acts that are still good (i.e. virtuous) but they do not go beyond the threshold. He presents it as a dilemma:

If the virtuous person has only one moral option in any particular situation, then the question of whether this choice is obligatory or supererogatory does not make sense. It is simply the only right choice. If, on the other hand, virtuous agents can choose between doing what is their duty and doing what goes beyond it, then there must be two kinds of virtuous people or virtuous choices. And indeed there are suggestions that we should distinguish between the 'maximally virtuous' and the 'minimally virtuous' person, between 'supreme' or 'saintly' virtue and ordinary or 'perfectionist' virtue. (Heyd 2015, p.32)

Heyd believes that virtue theory has no theoretical resources to make such a distinction: "But such categorical distinctions between degrees of virtue are not easy to explain."¹³⁹ Of course people can be more or less virtuous, but can there be moral models that are more or less virtuous? Heyd claims that this is impossible because the less virtuous "cannot themselves serve as a model for the non-virtuous individual trying to become a good person."¹⁴⁰ I must confess that I do not understand why the person giving 10\$ to charity cannot be a (modest) moral model, alongside the person who gives much more. Nevertheless, Heyd concludes that "Aristotle's doctrine of the mean seems to suggest that indeed virtue allows for no such categorical degrees."¹⁴¹ Aristotle's doctrine of the mean is a delicate and ample topic that I cannot hope to undertake here. However, I can hope to be able to respond to this argument by invoking arguments made by Rebecca Stangl (2015) against similar objections raised mainly by Crisp (2013). Her main goal is to give an account of supererogation in neo-Aristotelian terms (i.e. what

¹³⁹ Heyd (2015, p.32).

¹⁴⁰ Heyd (2015, p.33).

¹⁴¹ Heyd (2015, p.33).

Heyd claims to be impossible), but I cannot get into the theoretical details of the account. However, I think that her reply – to the claim that an Aristotelian account of virtues does not have the resources to distinguish between more virtuous and less virtuous or to establish a threshold for supererogation – is the correct reply to give.

Rebecca Stangl (2015) takes into account at least two possible interpretations of Aristotle's doctrine of the mean. The basic intuition behind Heyd's objection is that, if the main idea of virtuous behavior is to do what is 'fitting' in the circumstances (the right response, at the right time, for the right reasons), then it cannot be something that is more or less virtuous because there is no such thing as more or less 'fitting':

This seems to suggest that, in every situation, there is only one possible action that qualifies as virtuous. Any other action, moreover, will count as vicious. If this is right, there can be no action whose commission is virtuous but whose omission would not be vicious. (Stangl 2015, p.352)

Her reply to this objection is that Aristotle himself uses the metaphor of hitting a target¹⁴² when speaking about the right way to act. This means that doing the right thing can be seen as falling within a certain range. One can hit closer to the center of the target or further away, but the right behavior would still count as 'hitting the target'. Consequently, there can be degrees of being right or being virtuous:

(...) that the simile also illustrates nicely how one virtuous action might be more excellent than another virtuous action. In most real world cases, one need not hit the bull's-eye to hit the target, but hitting the bull's-eye might be the most praiseworthy shot of all. Likewise, hitting just shy of the bull's eye might be more praiseworthy than just making it onto the target. If hitting the mean is analogous to hitting a target, circumstances in which one can and does perform a supererogatory action might be analogous to hitting closer to the bull's-eye. (Stangl 2015, p.353)

I think this is a convincing reply. The metaphor is particularly apt in several other respects (I am grateful to Paul Woodruff for pointing this out to me). For example, it can leave

¹⁴² Stangl (2015) quotes Peter Losin (1987) as having this interpretation of Aristotle.

room for variations in results depending not only upon the agent, but also on luck: for both the marksman and the moral agent, wind and other contingencies might play some role in attaining a good result. Even more significant, there is a minimal qualification that one has to achieve in order to qualify as marksman and, similarly, as a virtuous agent. But once this minimal qualification is attained, there is always room for improvement because the perfect marksman and the perfectly virtuous are ideals. Therefore, it might well be the case that one action or one person is more virtuous than the other, because it can be closer or further away from the ideal – which shows that Heyd had no good reason to worry about this aspect. Most importantly, I think, the metaphor of the target reveals another way in which virtue theory can accommodate and explain supererogation: famously, omitting a supererogatory deed should be allowed (i.e. should bring no blame or punishment); similarly, one might say, missing the target does not mean that one did anything wrong, only that there is room for improvement, that one can do better.

However, I also believe that Heyd's skepticism towards a virtue-theoretical account of supererogation is not one rooted in the difficulty of the theoretical construction or the difficulties in interpreting Aristotle. I think that the uneasiness towards virtue theory comes from a more general worry, a worry that, broadly speaking, virtue theory cannot account for a strong notion of obligation, one that can serve as a firm 'threshold' of duty. This is a worry capable of raising the doubt that a firm distinction can be made between actions that go beyond the threshold and those which do not. For what it is worth, I think Heyd's worry is misguided and that virtue ethics is, indeed, the best candidate for accounting for supererogation.

Nevertheless, I believe that Heyd's way is an interesting way of not seeing things correctly, because it shows how the adherence to an obligation-centered frame can distort one's view about other theoretical frames. Heyd might seem right in his unspoken intuition that it is awkward to ask, in a virtue theory framework, the same kind of question one can easily ask in an obligation-centered framework, namely “how virtuous should someone be to be deemed virtuous

enough?” The question springs from the idea of a threshold, of course. The idea is that it is easy to conceive a threshold for obligation and to ask “when does obligation end, when has someone fulfilled their moral obligation?” but not so easy for virtue. As the discussion about utilitarianism and supererogation has pointed out, establishing the limit between obligation and supererogation can be seen in terms of admitting satisficing requirements instead of maximizing, i.e. about being required to do *enough instead of one's best*. Therefore, it would seem that the same question is justified in a virtue theory environment: “when would be virtue deemed as enough?”. The point of 'enough' being needed, of course, as a mean of establishing what is supererogatory and what is not. However, I think the the impression of strangeness comes not from some supposed intrinsic impossibility to come with an answer or a threshold, but from the suppositions hidden in the question itself. This is why, I think, Heyd's failure is relevant, because Heyd questions virtue theory while preserving strong presuppositions belonging to morality of law. The point in finding out when obligation stops or when there is enough virtue is that the agent looks to minimize own losses (which is a rational strategy under some hypotheses). The question “when is virtue enough?” makes sense only under a zero-sum game hypothesis *that acting virtuously means acting only for the general good, and acting for the general good implies acting against one's own self-interest*. Of course, this is a hypothesis that is false in a virtue theoretical environment.

Virtue theory, I believe, can establish a threshold for supererogation, but in doing so it will not answer this particular question, because this is a question that can be meaningfully asked under another set of premises.

4.10. Supererogation and Virtue: Conclusion

The previous chapters were dedicated to underlining the difference between various problems of supererogation. The present fourth chapter emphasizes their connection by bringing to light a particular view of morality, one present behind most conflicts represented by problems of supererogation. I have labeled this view of morality, “morality of law” because it conceives of moral rules in analogy with laws and it is centered upon the obligation to follow general, impersonal moral requirements. To remind, the main tenet in a morality of law is that if some deed is deemed to be morally good in an impersonal, impartial, objective manner, then that should be deemed morally obligatory, even if this demands considerable sacrifices from the agent.

It is clear why such a tenet would place morality of law at odds with supererogation: because it is an instance of the good-ought tie-up principle. More difficult to see was why would a theory adopt such a counter-intuitive supposition, namely that all morally good deeds should be made obligatory? My response is that there are two plausible sources for this supposition: first, it is the moral intuition that one *ought* to help whenever one is able to;¹⁴³ second, it is a certain theoretical framework (its historical context is described by MacIntyre, 1984), one that gives the blueprint for what I have called (following MacIntyre, 1984) “morality of law”. According to the morality of law framework, the general good diverges from the personal good and humans are seen as naturally egoistic. Consequently, according to this view, morality (seen as having to do with the public good) and self-interest are locked in a zero-sum game: whatever someone does for the public good it counts as a loss for the self and the other way around. Acting morally is seen as a difficult and demanding task and this is why, in order to counter-balance the (presupposed) naturally powerful self-interest, moral obligation is brought into picture to play the role of counterbalancing force. The result is that any morally good deed has to be subsumed

¹⁴³ This intuition alone is not enough for obtaining the good-ought tie-up, but it is a plausible starting point.

under the ought of moral obligation because of a permanent risk of agents ignoring the demands of morality to follow their own interest (morally good deeds are implying a cost that, one suspects, is not going to be voluntarily undertaken). Obligation becomes, therefore, the core of this moral view, of morality of law.

I have argued that wherever these presuppositions, of morality of law, are adopted, the adopting theory will not be a friendly environment for supererogation. But what would it be a friendly theoretical environment for supererogation? The second part of this last chapter is concerned with arguing in favor of the thesis that a frame pertaining to virtue theory is a theoretical environment that can accommodate supererogation with less difficulty.

I have offered two main reasons in favor of this thesis and I have also offered some arguments for why I think Heyd's claim (2015) that virtue theory cannot accommodate supererogation is wrong.

The main reasons in favor of virtue theory as supererogation-friendly, were constructed in the following manner: I have picked two presuppositions that are central to morality of law and which make accommodating supererogation difficult and then I have shown that removing or replacing them with suppositions of virtue theory¹⁴⁴ would lead to the removal of some problems or threats for supererogation.

In short, I have claimed that virtue theories are better candidates for accommodating supererogation because they are not obligation-centered and they do not have the supposition of a zero-sum game between the personal and the public good.

¹⁴⁴ I am aware of the wide variety of theories of virtue. The traits discussed are chosen to be as general as possible in order to belong to most of the theories in this group and at the same time specific only to this group of theories.

Chapter Five: Problems of Supererogation: Final Remarks and Future Study

The main ideas I have argued for in the previous chapters are, first, that what is usually called “the problem of supererogation” would be more accurately described as a family of problems (i.e. different but related problems); second, that the difference between problems is generated by the disruption brought by supererogation within different theoretical frames; and, third, that the various problems of supererogation are actually different points of conflict between supererogation and various traits of a certain general view about morality. Because in most of cases, supererogation clashes with the same general picture of morality, the result is that the problems, even though different, are connected among themselves because they represent conflicts with the same conception of morality (i.e. morality as obligation-centered, impersonal, impartial, and legalistic and difficult to follow)

The first chapter presents supererogation, its problems and how the discussion about supererogation started (by Urmson, 1958) by pointing out its odd status within morality, as the excellent moral deed that cannot be required. It also underlines supererogation's tendency to bring into discussion the status of moral obligation and the limits of the domain of morality in general.

The second chapter presents the actual differences between various problems of supererogation in consequentialist and utilitarian frameworks together with their proposed solutions. The third chapter is also a presentation of problems of supererogation, but in terms of reasons for action. In both chapters, the attempt to make room for supererogation usually resulted in introducing exceptions to general rules (e.g. agent-centered prerogatives), or introducing alternatives to impersonal/impartial evaluations of what is morally best (e.g. agent-relative

reasons vs. agent-neutral reasons) or allowing for restrictions of the scope of moral requirements (e.g. satisficing, agent-centered permissions).

The fourth chapter brings the various problems of supererogation together by showing that what they have in common is what supererogation is opposed to, namely the image of morality as a 'morality of law': obligation-centered, impersonal, impartial and requiring sacrifices from the agent. The chapter ends by pointing in the direction of a theoretical framework that might be more welcoming for supererogation, namely a 'morality of virtue' type of frame.

There are many issues that unavoidably remain untreated or insufficiently treated. The regret of having to let them be that way is only compensated by the hope that they might be future topics of study. I would like to describe briefly two of them.

First, I would like to refer to an issue related to the self-other asymmetry, as Slote has called it. Several times I was under the impression that some problems of supererogation were depending upon an equivocation between what one can require from oneself and what one can require from others. To remind, I have claimed that the expression "this is something we all should do" has this kind of ambiguity hidden in the usage of the pronoun "we". I think that this kind of asymmetry may be responsible also for the departure from the impersonal and impartial frames of thinking and that more could be said about supererogation by studying the consequences of a second person point of view in morality.

Second, I would like to refer to the issue that was treated only in passing at the end of the last chapter, namely the relation between virtue ethics and supererogation. It is obviously a topic larger than I can treat here. I cannot say, here and now, if adopting a virtue ethics frame is supposed to solve, dissolve or prevent some of the conflicts that are the core of problem proliferation regarding supererogation. However, I think one might have better chances to account for supererogation in a virtue ethics frame mainly because the problems of supererogation seem to have in common a certain puzzlement about *morality functioning without*

the compelling force of obligation. And this is a puzzle only when the presupposition is that the core of morality *is, actually, obligation.* In other words, it is puzzling that some morally excellent deeds are not obligatory simply because, generally speaking, what is morally good is also seen as obligatory (or as something one should do). I think virtue ethics, not having at its core this presupposition and not being obligation-based, has an obvious initial advantage. How this project might proceed from this point, is a matter for future research.

References:

- Crisp, R. (2013) Supererogation and Virtue. In: M. Timmons (ed.), *Oxford Studies in Normative Ethics* 3, pp. 13-34.
- Crisp, R. & Slote, M.A. (1997) *Virtue Ethics* (eds.). Oxford: Oxford University Press.
- Dancy, J. (1988) Supererogation and Moral Realism. In: J. Dancy, J.M.E. Moravcsik & C.C.W. Taylor (eds.), *Human Agency: Language, Duty, and Value: Philosophical Essays in Honor of J.O. Urmson* (pp. 170-188). Stanford, CA: Stanford University Press.
- Dancy, J. (1993) *Moral Reasons*. Oxford: Blackwell.
- Dancy, J. (2004) Enticing Reasons. In: R. Jay Wallace, Ph. Pettit, S. Scheffler & M. Smith (eds.), *Reason and Value: Themes From the Moral Philosophy of Joseph Raz* (pp. 91-118). Oxford: Clarendon Press.
- Dreier, J. (2004) Why Ethical Satisficing Makes Sense and Rational Satisficing Doesn't. In: M. Byron (ed.), *Satisficing and Maximizing* (pp. 131-154). Cambridge: Cambridge University Press.
- Foot, Ph. (1985) Utilitarianism and the Virtues. *Mind* 94, pp. 196-209.
- Hampton, J. (1992) Selflessness and the Loss of Self. *Social Philosophy and Policy* 10, pp. 135-165.
- Harman, E. (2016) Morally Permissible Moral Mistakes. *Ethics* 126, pp. 366-393.
- Heyd, D. (1982) *Supererogation: Its Status in Ethical Theory*. Cambridge: Cambridge University Press.

- Heyd, D. (2002/2016) Supererogation. In: Ed.N.Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Available at: <https://plato.stanford.edu/entries/supererogation/>
- Heyd, D. (2015) Can Virtue Ethics Account for Supererogation? *Royal Institute of Philosophy Supplement* 77, pp.25-47.
- Hurka, Th. (1990) Two Kinds of Satisficing. *Philosophical Studies* 59, pp. 107-111.
- Hurka, Th., & Schubert, E. (2012) Permissions to Do Less Than the Best: A Moving Band.F *Oxford Studies in Normative Ethics* 2, pp. 1–27.
- Jenkins, C.S., & Nolan, D. (2010) Maximising, Satisficing and Context. *Noûs* 44, pp. 451-468.
- Kagan, S. (1989) *The Limits of Morality*. Oxford: Clarendon.
- MacIntyre, A. (1984) *After Virtue: A Study in Moral Theory* (2nd ed.). University of Notre Dame Press.
- Mellema, Gr. (1991) *Beyond the Call of Duty: Supererogation, Obligation, and Offence*. State University of New York Press.
- McGoldrick, P.M. (1984) Saints and Heroes: A Plea for Supererogatory. *Philosophy* 230, pp. 523-528.
- Mill, J.S. (1861/2001) *Utilitarianism*. Hackett.
- Nagel, Th. (1986) *The View from Nowhere*. New York & Oxford: Oxford University Press.
- New, Chr. (1974) Saints, Heroes and Utilitarians. *Philosophy* 49, pp. 179-189.
- Nussbaum, M.C. (2004) Emotion as Judgments of Value and Importance. In: R.C. Solomon (ed.), *Thinking About Feeling: Contemporary Philosophers on Emotions* (pp. 183-199). Oxford: Oxford University Press.
- Portmore, D.W. (2003) Position-Relative Consequentialism, Agent-Centered Options, and Supererogation. *Ethics* 113, pp. 303-332.

- Pybus, E.M. (1982) Saints and Heroes. *Philosophy* 220, pp. 193-199.
- Pybus, E.M. (1986) A Plea for the Supererogatory: A Reply. *Philosophy* 238, pp. 526-531.
- Railton, P. (1984) Alienation, Consequentialism, and the Demands of Morality. *Philosophy and Public Affairs* 13, pp. 134-171.
- Railton, P. (2011) Two Cheers for Virtue: Or, Might Virtue Be Habit-Forming? *Oxford Studies in Normative Ethics* 1, pp. 295-330.
- Raz, J. (1975a) Permissions and Supererogation. *American Philosophical Quarterly* 12, pp. 161-168.
- Raz, J. (1975b/1999) *Practical Reason and Norms*. Oxford & New York: Oxford University Press.
- Rawls, J. (1971) *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Ross, W.D. (1930) *The Right and the Good*. Oxford: Clarendon.
- Scheffler, S. (1994) *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*. Oxford: Oxford University Press.
- Slote, M.A. (1984a) Morality and Self-Other Asymmetry. *Journal of Philosophy* 81, pp. 179-192.
- Slote, M.A. (1984b) Satisficing Consequentialism – I. *Proceedings of the Aristotelian Society* 58, pp. 139-163.
- Slote, M.A. (2004) Two Views of Satisficing. In: M. Byron (ed.), *Satisficing and Maximizing* (pp. 14-29). Cambridge: Cambridge University Press.
- Stangl, R. (2015) Neo-Aristotelian Supererogation. *Ethics* 126, pp. 339-365.
- Urmson, J.O. (1958) Saints and Heroes. In: A. Melden (ed.), *Essays in Moral Philosophy* (pp. 198-216). Seattle: University of Washington Press.

- Vessel, J.P. (2010) Supererogation for Utilitarianism. *American Philosophical Quarterly* 47, pp. 299-319.
- Williams, B.A.O. (1976) Persons, character and morality. In *Moral Luck: Philosophical Papers, 1973-1980*. Cambridge: Cambridge University Press.
- Williams, B.A.O. (1997) Morality, the Peculiar Institution. In: R. Crisp & M.A. Slote (eds.), *Virtue ethics* (pp.45-65). Oxford: Oxford University Press.
- Wolf, S. (1982) Moral Saints. *Journal of Philosophy* 79, pp. 419-439.
- Woodruff, P. (1991) Virtue ethics and the appeal to human nature. *Social Theory and Practice* 17, pp. 307-335.
- Woodruff, P. (2014) *Reverence: Renewing a Forgotten Virtue* (2nd ed.). New York & Oxford: Oxford University Press.
- Zimmerman, M.J. (1993) Supererogation and Doing the Best One Can. *American Philosophical Quarterly* 30, pp. 373-380.